THE PHYSICAL SOCIETY

Reports on

PROGRESS IN PHYSICS

VOLUME III

GENERAL EDITOR
ALLAN FERGUSON

Reprinted by
Wm. Dawson & Sons Ltd., London
with the Permission of the Publisher

Dutch reprint

CONTENTS

PAGE

General physics 1

§ 1. Surface tension. By R. C. BROWN

§ 2. Elasticity. By H. L. COX and D. G. SOPWITH

§ 3. Viscosity. By G. BARR

Fluid motion. By A. FAGE 26

The upper atmosphere. §§ 1 to 11 by S. CHAPMAN. §§ 12 to 16 by S. CHAPMAN and W. C. PRICE 42

Atomic physics. By N. FEATHER 66

The conservation of energy and momentum in elementary processes. By A. NUNN MAY 89

Sound. By E. G. RICHARDSON 106

The measurement of noise. By G. W. C. KAYE 130

Heat. By R. W. Powell 143

Magnetism. By L. F. BATES 175

Experimental electricity and magnetism 206

§ 1. Dielectrics. By L. HARTSHORN

§ 2. Alloys for permanent magnets. By D. A. OLIVER

§ 3. Soft magnetic materials and magnetic measurements. By C. E. WEBB

§ 4. Piezo-electricity. By P. VIGOUREUX

Electrical methods of counting. By C. E. WYNN-WILLIAMS 239

Superconductivity and the theory of metals. By A. H. WILSON 262

Photoelectricity. (Covering the years 1934–1935). By G. A. BOUTRY . . . 272

Optics. *Editor*, L. C. MARTIN 290

§ 1. General introduction. By the editor of the article

§ 2. Optical theory. By L. C. MARTIN

§ 3. Optical instruments (general). By L. C. MARTIN and B. K. JOHNSON

§ 4. Physical optics. By F. SIMEON

§ 5. Ophthalmic optics. By WILLIAM SWAINE

§ 6. Colour-vision. By W. D. WRIGHT

§ 7. Glare. By W. S. STILES

§ 8. New light-sources: recent developments in electric discharge lamps. By J. T. RANDALL

PAGE

X rays. By THE X-RAY STAFF, PHYSICS DEPARTMENT, NATIONAL PHYSICAL LABORATORY. *Editor*, G. W. C. KAYE 338

§ 1. Production of X rays. By G. E. BELL

§ 2. Absorption of X rays. By I. BACKHURST

§ 3. Scattering of X rays. By I. BACKHURST.

§ 4. Ionization in gases. By I. BACKHURST

§ 5. The energy expended in ionizing air. By W. BINKS

§ 6. Absorption spectra. By G. SHEARER

§ 7. Crystal structure of metals and inorganic substances. By W. A. WOOD

§ 8. Crystal structure of organic and biological substances. By J. THEWLIS

§ 9. Radium. W. E. T. PERRY

Spectroscopy. *Editor*, W. JEVONS 369

§ 1. Atomic spectra. By M. F. CRAWFORD

§ 2. Astronomical spectroscopy. By A. HUNTER

§ 3. The application of spectroscopy to photochemistry. By R. W. B. PEARSE

§ 4. The rotation-vibration spectra of the lighter hydrocarbons. By W. H. J. CHILDS

GENERAL PHYSICS

§ 1. SURFACE TENSION

By R. C. Brown, Ph.D.

It is possible in this article to notice only a proportion of those papers dealing with surface tension which have appeared during the last year or so. Among those omitted is a considerable number of papers descriptive of the use of well-known methods for the purpose of determining more accurate data for the surface tensions of pure substances and mixtures. Such work is of great value; but the most fitting place for its recognition is the future edition of a table of constants, and it is to be hoped that the results of all painstaking measurements will be included in such volumes. The supply of reliable data for even the simplest properties of pure substances and mixtures is notoriously scanty.

The surface tension of water. Dynamic and static methods. The first subject which must be mentioned is the determination of the surface tension of water. This may surprise the non-specialist reader, but the fact is that it is only during the past twenty years that even as few as three or four methods of measuring surface tension have been made to yield concordant results for water. The credit for most of the work done so far lies with the Americans. In the case of capillary rise[1,2,3] experimental technique has been perfected and uncertainties in the theory hardly exist when narrow tubes are used. For the drop-weight[4] and ring[5] methods tables of experimental corrections have been constructed. When using any of these three methods, future workers, particularly those in the field of industrial research, would do well to consult the original papers to which reference has been made.

But there are other methods and their number is increasing. And to make matters worse there is a persistent belief that the tension of a water surface does not reach a steady value until a hundredth of a second or so after its formation. Consequently "dynamic" methods, i.e. those in which the surface is moving when measurements are taken, are sometimes suspected of yielding high values. Quincke was convinced of the reality of this effect and Lenard's school at Heidelberg claim to have established it beyond doubt. In recent years the question has been studied by Hiss[6], Schmidt and Steyer[7], Kleinmann[8], and Seitz[9], all of whom have used a capillary rise method. By blowing a jet of air across the top of the tube the meniscus was drawn up and kept fresh, and its rate of fall was then observed after the air stream was shut off. Allowance must, of course, be made for the finite rate of fall which would be observed even if the surface tension were constant. Schmidt and Steyer found an exponential fall of surface tension from an initial value of 96·6 dynes per cm. through 87·5 at one thousandth of a second to 74·0 at thirteen thousandths of a second. The experimental difficulties must be great.

Bohr[10, 11], whose work on the Rayleigh oscillating jet method must be regarded as authoritative, found a value of 73·2 dynes per cm. at 12° C., although Rayleigh's own work on jets of dilute sodium oleate solutions which gave pure water values showed that the surface was renewed. But Bohr's surface was six hundredths of a second old and so, according to the results of Schmidt and Steyer, had reached maturity. On the other hand Pedersen[12], using the jet method a year or so before Bohr, found a value of 74·3 dynes per cm. at 15° C. and estimated the age of his surface at less than one hundredth of a second.

The reason for going over this ground in an article mainly concerned with 1935 is that the question is raised again by several papers under review.

The first, that of W. N. Bond[13], was noticed in last year's *Reports*. Two vertical collinear water jets produce a horizontal disc after collision. Measurement of the disc and the rates of flow of the jets leads to a value for water of 73·8 dynes per cm. at 15° C. This is in fair agreement with accepted *static* values.

But another purely dynamic method described by Buchwald and König[14] gives a result of 81·5 dynes per cm. at 12° C. The experiment consists of forming "water bells" such as were studied by Savart[15]. A water jet directed vertically downwards on to the centre of a horizontal circular metal disc forms a "bell" of which striking photographs are given in the original paper. The theory of the phenomenon was worked out by Boussinesq[16, 17], and need not be reproduced here.

The above-mentioned value (81·5 dynes per cm.) is the mean of a number of fairly consistent results, the age of the surface varying from 0·029 to 0·078 sec.—always in excess of the time required to reach equilibrium as determined by Schmidt and Steyer. Buchwald and König attempt no very definite explanation of their result but suggest that the static surface actually corresponds to an equilibrium between the various single and associated water molecules and that this is prevented from being established in their experiments. It will be interesting to see what results they obtain after using, as they suggest, a liquid which is known to be unassociated.*

Apart from the surface-tension values which they yield the experiments are worthy of note on account of their beauty. A truly bell-shaped surface is formed when the water is allowed to fall on to a plane water surface instead of uniting on the axis of the bell.

The surface tension of water has also been determined by Prosad and Ghosh[18] who have repeated the experiments of Maass[19] without using the full theory quoted by him. When a vertical jet of liquid falls on to a horizontal plane, standing waves are set up on its surface, the principle being the same as that underlying the formation of a standing pattern on a stream flowing past an obstacle. Measurements of wave-length, diameter of jet (both of which increase with distance from the obstacle upwards) and the total discharge of the jet lead to a value for the surface

* The results of these experiments have actually been published[45]. Buchwald and König find that for benzene the dynamic and static values are identical; but for ethyl alcohol, acetic acid, ethyl acetate and carbon tetrachloride the dynamic surface tension is anything from 5 to 16 per cent higher than the static surface tension.

tension. The authors are not concerned with the "dynamic-versus-static" controversy and their results give $\gamma = 71{\cdot}6$ dynes per cm. at 20° C., their wave-lengths being of the order of a millimetre or so.

Satterly and McPherson[20] performed similar experiments in 1934 but with little success as far as the consistency of the results was concerned, especially in the case of mercury. Now Satterly and Strachan[21] have published some further results. They allowed their jet to fall on an inverted round-bottomed flask. The troughs and crests of the standing pattern on the jet acted as curved mirrors and produced point images when illuminated by parallel light. These enabled the wave-length to be measured quite accurately. The surface tension of water ranged from 70·5 to 76·0 at 15° C., wave-lengths being of the same order as those of Prosad and Ghosh. Clearly no small amount of work has yet to be done on the so-called "dynamic" methods of measuring surface tension.

It may be of interest to note that the author[22] of this article showed that in the case of the ripple method, which is sometimes looked upon as giving fresh surfaces, the attainment of surface equilibrium of a slowly adsorbed solution was in no way hindered by the passage of the waves over the surface. The rate of fall of surface tension with time was, if anything, greater than that found by Dr N. K. Adam[23] who measured the same solution by a static method. Thus as far as the degree of freshness of the surface is concerned the ripple method is in no way comparable to that of oscillating jets.

The surface tension of heavy water has been measured by several workers. Selwood and Frost[24] used several concentrations, the highest being 92 per cent and this gave 68·1 dynes per cm. By extrapolation they estimated the surface tension of pure D_2O at 67·8 dynes per cm. They do not mention which method of measurement they adopted. Lachs and Minkow[25] made careful capillary rise experiments and give $72{\cdot}2 \pm 0{\cdot}5$ dynes per cm. for 99 per cent. Flood and Tronstad[26] using a bubble-pressure method find that the surface tension of D_2O is slightly higher than that of H_2O but the difference is very little greater than their experimental error. It persists, however, at 50° C. where the value for the heavy water is 67·7 dynes per cm.

The surface tension of mercury. Considerable advances have been made in the determination of the surface tension of mercury, and among recent work is a noteworthy paper by H. Brown[27], who has measured by both the drop-weight and the sessile drop methods the surface tension of mercury *in vacuo* as well as in contact with dry air, saturated water vapour and liquid water. As Brown remarks, there is chaotic disagreement between the results of former workers, and his experiments are designed to investigate the influence of air and water vapour on the mercury surface with a view to establishing standard conditions. The drop-weight results were reduced by Harkins and Brown's[4] method and those for the sessile drop from Worthington's formula or by Porter's method of choosing a drop of such a size as to make the correction vanish.

The results as found by the two methods both for mercury *in vacuo* and against water agree to 0·3 per cent, the values being 472 and 374 dynes per cm. respectively.

When the sessile drop method was used with mercury in dry air the surface tension fell but slowly with time. A more rapid fall was noticed when the air was moist; but when water vapour alone was used there was very little reduction of surface tension unless the vapour was saturated, when the value fell from 447 to 430 dynes per cm. in a day. In a mixture of saturated vapour and air the initial value was the same but the fall was greater. With the drop-weight method the fall with time was not so noticeable.

The author suggests that his results may be explained on the hypothesis that the mercury surface is almost unaffected by dry air but is oxidized in the presence of moist air. This may explain the diversity of results hitherto published by other writers. Brown's results for a stable surface are the same by both methods and agree well with those obtained by several experimenters who used the maximum bubble-pressure method.

It is worthy of note that Glidden[28], using the drop-weight method to measure the surface tension of a mercury-mercurous sulphate solution interface and making use of Harkins and Brown's corrections, finds a value for the mercury-water interface almost identical with that of Brown.

It would seem that we now have really authoritative values for the mercury surface *in vacuo* and against water, at any rate when static* methods are used.

This reservation is rendered necessary by the results of Satterly and Strachan[21] whose work on water jets has been mentioned above. These authors have used the same jet method for the surface tension of mercury in air, the obstacle on which the jet impinged being a plane mercury surface. Their results lie between 548 and 577 with a mean value of 563 dynes per cm. for five different jets. When the mercury was allowed to stand in dishes exposed to the air the figure fell to 499 in three weeks and the authors are of the opinion that it was approaching asymptotically to 495 dynes per cm.

These results are rather disturbing and Satterly and Strachan quote a paper by Popesco[29] which has been quite undeservedly ignored by a good many writers. This author made an exhaustive study of the sessile drop method as applied to mercury, and found that when the surface was formed in a gas the tension was perhaps 100 dynes per cm. higher than the value found when the surface was formed in a vacuum. He suggests that the gas molecules hinder the arrangement of the mercury molecules into positions of minimum potential energy. Satterly and Strachan draw attention to a flaw in the jet method namely the possibility that the velocity at the surface may be less than the average velocity of the jet which is the one calculated from measurements of the total discharge.

The calculation of surface tension from experiment. Prof. Porter has continued his valuable work, one of his contributions being a paper[30] in which curves are drawn enabling the capillary rise experimenter to deduce surface tensions no

* The author of this article is not persuaded that the drop-weight and maximum bubble-pressure methods are truly static, but neither are they dynamic in the sense that the surface is *continually* renewed as it is in jet experiments. Thus failing the use of terms such as "quasi" and "pseudo" it would seem best to reserve "dynamic" for the jets and kindred experiments and to label all other methods "static".

matter what may be the size of his tube. The angle of contact must, however, be zero.*

Using Prof. Porter's non-dimensional quantities we may write

$$\frac{\beta^2}{r^2}=\frac{1}{2}\frac{h_0}{r}+\Delta,$$

where $\beta^2=\gamma/g\rho$, r is the radius of the tube, h_0 is the elevation of the vertex of the meniscus, and Δ is the correction.

Basing his calculations on the tables of Bashforth and Adams and on Rayleigh's equation Prof. Porter has graphed Δ against $\frac{1}{2}h_0/r$.

The second part of the paper deals with the height, h, of the *meniscus* in wide or narrow tubes. This quantity is also the height of the vertex of a sessile drop above its horizontal section of maximum radius. In a previous paper[31] noted in the first volume of these *Reports* Prof. Porter has written

$$\frac{\beta^2}{r^2}=\frac{1}{2}\frac{h^2}{r^2}-\Delta$$

and has plotted Δ against h^2/r^2 up to $h^2/r^2=0{\cdot}25$. When this value is exceeded Δ becomes negative and finally infinite. In order to avoid this he now writes

$$\frac{\beta^2}{r^2}=\frac{\frac{1}{2}h^2/r^2}{(1-h^2/r^2)+\Delta'}$$

and plots Δ' against h/r up to $h/r=1$, which is, of course, the upper limit of its value when the drop is so small as to be spherical or when the capillary tube is exceedingly narrow. Thus surface tensions may now be calculated from sessile drop (or bubble) experiments no matter what the size of the drop, although very little accuracy may be expected when small, almost spherical, drops are used, since the principle of the experiment is a comparison between the effects of gravity and surface tension on the shape of the surface.

In another paper[32] Prof. Porter reverts to a question which he has raised previously, namely that, whatever may be the radius of a tube, it will behave as a wide one at temperatures approaching the critical temperature of the liquid, i.e. when the surface tension becomes vanishingly small. Thus under these conditions the effect of capillarity may be merely to produce a curved meniscus but no measurable rise of its vertex, and the calculation of surface tension then becomes a matter of evaluating the correction appropriate to an *infinitely wide* tube. Hence the critical temperature must not be looked upon as that temperature at which capillary rise becomes immeasurably small since this will depend on the radius of the tube used. Prof. Porter hints that the fact that observations on a water meniscus yield a value of 374° C. for its critical temperature while density measurements place it some 6° C. higher may find its explanation here.

The Eötvös equation. The simple Eötvös relation may be written

$$\gamma\left(\frac{M}{\rho}\right)^{\frac{2}{3}}=K\,(T_c-T) \qquad \ldots\ldots(1),$$

* Prof. Porter has now (Oct. 1936) treated the case of finite contact angle[46].

and it is well known that apart from a certain number of substances K varies from liquid to liquid and with temperature.

Now the factor $(M/\rho)^{\frac{2}{3}}$ represents an estimate of the area which a single molecule occupies in the surface and Dervichian[33] suggests that we should replace this by the appropriate area which is deduced from mono-molecular film experiments (e.g. 24 A^2 for the fatty acids, 25 A^2 for some alcohols and so on). He finds then that the value of K in all these cases becomes about 1·2 even though the original Eötvös constant varies between 0·9 and 4·2. Other examples are given. Actually two groups emmerge, one, consisting of liquid elements, for which $K=2$ approximately, and another in which $K=1$, embracing those liquids which have unsymmetrical polar molecules.

Dervichian has also published a paper[34] in which he calculates surface energy by adopting the well-known principle of the breaking of a liquid column with the production of two new surfaces. He obtains the equation

$$\gamma u^{\frac{2}{3}} = \frac{2{\cdot}3a}{6MN^{\frac{1}{3}}} \cdot \rho \log_{10} \frac{\rho'}{\rho} \qquad \text{.....(2),}$$

where u is molar volume M/ρ, a the constant in the van der Waals equation, M molecular weight, N Avogadro's number, and ρ, ρ' the density of liquid and vapour respectively.

By consulting tables Dervichian has found the empirical relation

$$\rho \log_{10} \frac{\rho'}{\rho} = \frac{C}{T_c}(T_c - T \pm \epsilon) \qquad \text{.....(3),}$$

where C is a different constant for each liquid and ϵ is of the order of a few degrees.

Thus, comparing equations (2) and (3) with equation (1) or, better still, with the Ramsay and Shields modification of the Eötvös equation, we see that the value of K is

$$\frac{2{\cdot}3a}{6MN^{\frac{1}{3}}} \cdot \frac{C}{T_c}.$$

Agreement with experimental values is moderate, the greatest deviation being with polar liquids. This is only to be expected since the use of the quantity $(M/\rho)^{\frac{2}{3}}$ to represent the molar area implies the assumption that the molecules are cubical in shape.

Miscellaneous experimental work. Two descriptions of micro methods of measuring surface tension have appeared. Each measures the force acting on a column of liquid the ends of which are situated in capillaries of different radii. Neither introduces any fundamentally new principle, and for details the original papers should be consulted. The authors are, in the one case, Natelson and Pearl[35], and, in the other, Mahajan[36]. The latter's method gives density as well as surface tension.

A novel experiment has been described by Pfund and Greenfield[37]. An indentation is produced in a liquid surface by directing on to it a vertical jet of air. When half a dozen liquids of known surface tension were used, it was found that a graph of surface tension against the reciprocal of the square root of the depth of the

depression was a straight line, and consequently the method could be used to determine surface tensions, particularly of viscous liquids. The authors note (without comment) that the densities of the various liquids do not enter into this empirical relation. Under certain conditions the jet partook of the nature of a sensitive flame, the indentation disappearing to the accompaniment of the classic bunch of keys.

Pietenpol[38] has used a simple bubble method for measuring the surface tension of molten glass. A bulb is blown and supported on a plaster of Paris collar in a furnace, the pressure inside the bulb being recorded on a manometer. As the temperature is raised the pressure and the diameter of the bulb are measured. The usual simple formula gives the surface tension of the glass if allowance is made for the weight of the glass in the top half of the bulb. With rising temperature the surface tension shows a sharp rate of fall until the melting point is reached and above that a linear graph is recorded. The values at the melting point range from 230 to 360 dynes per cm. according to the composition of the glass.

Léauté[39] has demonstrated that capillary rise may be used with very viscous liquids, namely tars and bitumens; in fact he is of the opinion that this method is more appropriate than that of maximum bubble pressure. The greatest time needed for the establishment of the true height was 360 hours.

Some original work on contact angles has been done by Mack[40]. If a very small drop of liquid is placed on a solid with which it forms an *acute* contact angle its shape may be taken as that of a spherical cap to a first approximation, and the contact angle may be expressed in terms of the height of the drop and the radius of its base. The height is difficult to measure but with a little algebraic manipulation it may be expressed in terms of the volume of the drop. Thus it is only necessary to form a drop from a calibrated tube and measure the radius of its base. For water on apple wax the contact angle was found to vary irregularly between about 50° and 90° as the radius of the drop decreased from 0·156 to 0·060 cm. After that, down to a radius of 0·039 cm., the angle was constant between 88° and 90°. By comparison with other methods it is shown that this angle is the *advancing* contact angle. Moreover in the case of small drops the angle cannot be altered by mechanical disturbance as it can with larger drops.

In another paper Mack and Dorothy Lee[41] consider drops forming obtuse angles of contact and have constructed tables, based on those of Bashforth and Adams, which enable the angles to be calculated from the dimensions of the drop.

Mlle Paulette Berthier[42, 43] has investigated the rise of liquids in strips of filter paper and her results show that it does not depend on surface tension. Dilute aqueous solutions, all of which had approximately equal densities and viscosities, were used, the solutes being chosen so as to give surface tensions ranging from 27 to 61 dynes per cm. The rise observed after 3 hours was the same in all cases provided the system was surrounded by air saturated with water vapour. Decharme in 1874 noted that a large number of solutions rose more quickly than water and attained greater final heights; but Mlle Berthier, the height of whose strips was 38 cm., finds no upper limit in the presence of saturated vapour. On the other hand Decharme's results were reproduced when the experiments were performed in

unsaturated air. When dry air was used the equilibrium heights were found to be greater for liquids with lower vapour pressure.

An exhaustive study of the surface tension of urine has been made by Perryman and Selous[44]. They find an exponential decrease of surface tension with time and a variation throughout the day of the values for fresh specimens. Oddly enough this variation is independent of the times of intake of food and drink. The main conclusion reached is that there is no relation between surface tension and any specific disease or group of diseases, and even if such a connexion did exist, diagnosis would be extremely difficult on account of the daily changes and the variation with time of a particular specimen.

References

(1) Richards and Coombs. *J. Amer. chem. Soc.* **37**, 1656 (1915).
(2) Richards and Carver. *J. Amer. chem. Soc.* **43**, 827 (1921).
(3) Harkins and Brown. *J. Amer. chem. Soc.* **41**, 503 (1919).
(4) Harkins and Brown. *J. Amer. chem. Soc.* **41**, 499 (1919).
(5) Harkins, Young and Cheng. *Science*, **64**, 333 (1926).
(6) Hiss. Diss. Heidelberg (1913).
(7) Schmidt and Steyer. *Ann. Phys.*, Lpz., **79**, 442 (1926).
(8) Kleinmann. *Ann. Phys.*, Lpz., **80**, 245 (1926).
(9) Seitz. *Ann. Phys.*, Lpz., **1**, 1099 (1929).
(10) Bohr. *Philos. Trans.* A, **209**, 281 (1909).
(11) Bohr. *Proc. roy. Soc.* A, **84**, 395 (1910).
(12) Pedersen. *Philos. Trans.* A, **207**, 341 (1907).
(13) Bond. *Proc. phys. Soc.* **47**, 549 (1935).
(14) Buchwald and König. *Ann. Phys.*, Lpz., **23**, 557 (1935).
(15) Savart. *Ann. Chim.* (*Phys.*), **54**, 55, 113 (1833).
(16) Boussinesq. *C.R. Acad. Sci.*, Paris, **69**, 45, 128 (1869).
(17) Boussinesq. *C.R. Acad. Sci.*, Paris, **157**, 89 (1913).
(18) Prosad and Ghosh. *Indian J. Phys.* **10**, 31 (1936).
(19) Maass. *Trans. roy. Soc. Can.* **9**, 133 (1915).
(20) Satterly and McPherson. *Trans. roy. Soc. Can.* **28**, III, 177 (1934).
(21) Satterly and Strachan. *Trans. roy. Soc. Can.* **29**, III, 155 (1935).
(22) Brown, R. C. *Proc. phys. Soc.* **48**, 312 (1936).
(23) Adam. *Trans. Faraday Soc.* **31**, 204 (1935).
(24) Selwood and Frost. *J. Amer. chem. Soc.* **55**, 4335 (1933).
(25) Lachs and Minkow. *Nature*, Lond., **136**, 186 (1935).
(26) Flood and Tronstad. *Z. phys. Chem.* **175**, 347 (1936).
(27) Brown, H. *J. Amer. chem. Soc.* **56**, 2564 (1934).
(28) Glidden. *J. Amer. chem. Soc.* **57**, 236 (1935).
(29) Popesco. *Ann. Phys.* **3**, 402 (1925).
(30) Porter. *Phil. Mag.* **21**, 703 (1936).
(31) Porter. *Phil. Mag.* **15**, 163 (1933).
(32) Porter. *Phil. Mag.* **20**, 1163 (1935).
(33) Dervichian. *C.R. Acad. Sci.*, Paris, **200**, 2065 (1935).
(34) Dervichian. *C.R. Acad. Sci.*, Paris, **201**, 333 (1935).
(35) Natelson and Pearl. *J. Amer. chem. Soc.* **57**, 1520 (1935).
(36) Mahajan. *J. sci. Instrum.* **13**, 189 (1936).
(37) Pfund and Greenfield. *Industr. Engng Chem.* (Analytical Ed.), **8**, 81 (1936).
(38) Pietenpol. *Physics*, **7**, 26 (1936).
(39) Léauté. *C.R. Acad. Sci.*, Paris, **201**, 41 (1935).
(40) Mack. *J. phys. Chem.* **40**, 159 (1936).
(41) Mack and Lee, Dorothy. *J. phys. Chem.* **40**, 169 (1936).

(42) BERTHIER, PAULETTE. *C.R. Acad. Sci.*, Paris, **199**, 1221 (1934).
(43) BERTHIER, PAULETTE. *C.R. Acad. Sci.*, Paris, **200**, 1105 (1935).
(44) PERRYMAN and SELOUS. *J. Physiol.* **85**, 128 (1935).
(45) BACHWALD and KÖNIG. *Ann. Phys.*, Lpz., **26**, 659 (1936).
(46) PORTER. *Phil. Mag.* **22**, 726 (1936).

§ 2. ELASTICITY

By H. L. COX, B.A. and D. G. SOPWITH, B.Sc.Tech., A.M.I.Mech.E.

The chief characteristic property of a solid is undoubtedly its ability to transmit shearing force. Indeed it is this property that renders the material solid and makes it usable for constructional purposes. The actual mechanism by which a solid body is able to transmit shearing force is not yet well understood, although for materials of ionic lattice type a fairly complete representation of the mechanism has been developed by Lennard-Jones and others; but experience has shown that the elastic behaviour of a large class of practical materials is governed by the simple empirical relation known as Hooke's law. Hooke's law in its generalized form states that each point of the stressed solid the states of stress and strain are related by a system of linear equations, and in general, it is the concern of the theory of elasticity to apply this empirical law to determine the states of stress and strain set up in a body of a given shape by the application of known forces or displacements. Although, in view of the atomic nature of actual materials, the conception of stress as the ratio of a load to an area round a point as the area is indefinitely reduced cannot be regarded as fundamentally satisfactory, nevertheless, the theory of elasticity founded upon this conception does enable the deformation and the strength of solid bodies of different shapes under different load systems to be correlated.

As a result of the definite and simple nature of Hooke's law, the differential equations governing any particular class of problem in elasticity may usually be derived fairly easily, and progress in the development of the subject largely consists in the derivation of solutions of known equations appropriate to certain boundary conditions. By one method, a solution satisfying to any required degree of approximation given boundary conditions may be developed by composition of a number of exact solutions of the differential equations; by another, using the principle of minimum energy (Rayleigh), a solution approximately satisfying the equations of equilibrium may be compounded from a number of assumed solutions all satisfying the given boundary conditions. The latter method proves particularly useful in problems of stability and in problems of vibration (of which stability problems are a particular case). At the same time, a certain amount of work is published from time to time dealing with the more general aspects of the theory. In the present review, it is proposed to deal with the latter type of work first, to continue by discussion of new solutions, grouped mainly according to the method of solution, and to conclude by a survey of recent developments in the theory of frameworks.

In usual applications of the theory of elasticity, it is assumed that the strains are everywhere small, and that they may be represented with sufficient accuracy by the first powers of the first differential coefficients of the displacements. If the permis-

sible elastic strain is not negligible in comparison with unity, or if the form of the body is such as to permit large displacements without correspondingly large strains, this assumption is not justified. For application to this type of problem, Seth[1] has developed a theory of finite strain. The method is based on the specification of the deformed shape of the body and the author emphasizes the importance of referring the components of strain and the body stress equations to the strained and not to the unstrained configuration of the body. As examples of the application of the method the author considers the problem of the extension of a cylinder under axial tension when the permissible elastic strain is large, the problem of a thin flat plate bent into the form of a portion of a circularly cylindrical surface, and the problem of a right circular cylinder subjected to torsion about its axis.

Although the fundamental basis of the theory of elasticity, Hooke's law, is very simple in form, application of the law is complicated by the number of variables involved. In general, there are six components of stress and six components of strain which are related by six linear equations. In addition, the condition that each element of the strained body must be in equilibrium gives three relations between the first differentials of the stress components and the condition that the displacement must everywhere be single-valued gives three relations between the second differentials of the strain components. Formally these equations enable all but one of the stress or strain components to be eliminated, leaving a differential equation involving only this one component. In practice it is often convenient to express the stress components by stress functions chosen in such a way as to satisfy the equations of equilibrium identically. Thus in the case of plane stress, the equations of equilibrium

$$\frac{\partial \widehat{xx}}{\partial x}+\frac{\partial \widehat{xy}}{\partial y}=0$$

and

$$\frac{\partial \widehat{xy}}{\partial x}+\frac{\partial \widehat{yy}}{\partial y}=0$$

are satisfied identically, if for the stress components $\widehat{xx}$, etc., we write

$$\widehat{xx}=\frac{\partial^2\chi}{\partial y^2},\quad \widehat{yy}=\frac{\partial^2\chi}{\partial x^2} \text{ and } \widehat{xy}=-\frac{\partial^2\chi}{\partial x\partial y}.$$

The function χ is known as Airy's stress function. Phillips[2] has investigated the physical meaning of this function and other functions of similar type. He has shown that in general such functions represent the moment about some fixed axis of all stresses acting across a surface bounded by one fixed line and a parallel line through the variable point.

The stresses set up in a body by the application of known forces may sometimes conveniently be determined by use of the principle that the distribution of stress must be such that the variation of the total energy (including the work done by the external loads) must be zero for any small displacement of the system from its equilibrium position. In the usual application of this principle, the strain energy is expressed as a function of the applied loads; but Bateman[3] has shown that a similar method is available when the deformation is caused by known displacements

instead of, or in addition to, known external forces, provided the strain energy can also be expressed in terms of the unknown forces.

Many problems, both in elasticity and in other branches of physics, require the solution of Laplace's equation for given boundary conditions. Seth[4] has shown that, if the problem is a two-dimensional one and if the boundary consists entirely of straight lines, a very general solution may be obtained by application of the analytical device of conformal representation. By this device the polygonal boundary is transformed into a straight line, the apices of the polygon being represented by a series of points along this line. Formally the method consists in the solution of the problem for the straight line boundary and subsequent conversion of this solution into the original coordinate system, the singularities in the neighbourhood of the apices of the polygon being explored by an independent method; but a complete explicit solution is not in general possible, the final conversion into the original coordinate system being impracticable. Seth[5] and Okubo[6] have applied the method to certain torsion problems.

The problem of the stress distribution set up in a cylinder of any given cross section by the application of torsion, shear or bending load is one of very wide practical interest. Solutions of the torsion and bending problems appropriate to certain forms of cross section were first published by St Venant; but discussion of the problem and related matters still forms the subject of many papers. Sunatani[7] has shown how the equations appropriate to both the bending and torsion problems may be derived simultaneously, and in collaboration with Minagawa[8] he has investigated the maximum stress in short beams in cases where the stress due to direct shear is high. Deimel[9] has investigated the effect of special end conditions on St Venant's solution of the problem of torsion of a circular cylinder and he has concluded that variations of the manner in which the torque is applied are likely to have very little effect.

The flexural centre of the cross section of a cantilever is defined as the point through which a force may be applied in any direction without causing the beam to twist. The precise meaning of twist in this connexion is not always easy to define and the position of the flexural centre is always to some extent dependent upon the precise conditions of fixity at the root of the cantilever. General methods for the determination of the position of the flexural axis of a non-tapered beam have been developed by Schwalbe[10], by Trefftz[11] and by Williams[12]. For symmetrical sections Schwalbe gives formulae by which the distance of the flexural axis from the centroid of section can be calculated, the method also being applicable to non-symmetrical sections, whilst Trefftz shows that the flexural axis defined by $x=0$, $y=0$ must be such that

$$\iint (t_x T_x + t_y T_y)\, dxdy = 0,$$

where t_x and t_y are the shear stresses due to torsion and T_x and T_y are the shear stresses due to direct shear. Williams has developed a method for finding the flexural axis of thin-walled sections and has shown that the flexural axis of an open section often lies completely outside the section.

The distribution of shear stress in a cylinder subjected to torsion is determined

by Laplace's equation; the same equation also governs the deformation of a membrane subjected to uniform pressure, the analogue to the shear stress in torsion being the slope of the membrane. Accordingly the distribution of shear stress in a cylinder of any section may be investigated experimentally by observation of the slope of a membrane stretched over a hole representing the contour of the section of the cylinder. This experimental method was first employed by Prandtl and by Taylor and Griffiths; but an improvement of the method particularly applicable to re-entrant sections has recently been devised by Föppl[13].

Another experimental method for the determination or verification of stress analyses is available for problems of plane stress, i.e. problems in which only three stress components need be considered. This method, known as the photo-elastic method, is based on the manner in which the optical properties of some materials vary in sympathy with the stresses to which they are subjected. This method, which has been developed and applied to a wide range of problems by Coker and Filon[14], has recently been used by Tuzi and Nisida[15] to investigate the effect of impact loads on beams and rings. Using a special high speed camera, Tuzi and Nisida found that the frequencies of oscillation and the times of contact were in good agreement with theoretical results; but that the actual maximum stresses were 10–20 per cent less than the theoretical values. Gough, Cox and Sopwith[16] have recently pointed out that, in certain cases, a partial check on theoretical stress analysis may be obtained by means of fatigue tests. Many materials fail in fatigue without previously undergoing any considerable plastic deformation, so that components may be fractured without any appreciable distortion from the original form. The position of the fracture in such components therefore indicates the point of maximum stress and the value of the stress range at this point calculated theoretically from the range of load applied may be compared with the normal fatigue range of the material. This method has been used with some success to check theoretical conclusions as to the stress distribution set up in the ring of an eyebolt.

Another class of problem, of which numerous theoretical investigations have been made, is that of the stresses in thin plates of various shapes supported and loaded in different ways. The differential equations governing such a problem depend upon the type of loading applied, whilst the boundary conditions relate to the shape of the plate and the manner in which the plate is loaded or supported at the edges. Some special problems of this type have recently been considered by Macgregor[17], whilst Habel[18] has treated the problem of a flat rectangular plate simply supported at the edges and loaded along a strip or strips parallel to one pair of edges. This problem represents an elastic floor supporting a party wall. In problems of flat plates, it is necessary to state two edge conditions in terms of the values of the four variables, the deflection, the slope (inclination normal to the edge), the bending moment and the shearing force. In many cases these two conditions are merely that two of these four variables should be zero at all edges; but Bateman[19] has recently considered a case in which one edge condition takes the form of a relation between the shear force and the deflection at the edge. This case is that of a rectangular plate subjected to uniform normal pressure and supported on flexible edge

members. The problem is solved by the composition of exact solutions, the edge conditions being satisfied by a process of successive approximation. Vint and Elgood[20] have considered another problem involving elastic support of a rather different type. In this problem, the edges of the plate are free from load and the reaction to a central load is that set up by the pressure of the plate on an elastic support with which it is in contact over the whole of its under surface. For the case of the circular plate, Vint and Elgood obtain an exact solution in Bessel series, whilst for the rectangular plate their solution is in the form of a double series of special functions, the coefficients of which are defined by a doubly infinite series of differential equations, solved by a process of successive approximation. Vint and Elgood have checked the theoretical values of the deflections by experimental methods and have obtained good agreement. Other types of special loading over the area of a circular plate with supported edges have been considered by Sen[21]. He has obtained exact solutions for symmetrical and asymmetrical load distributions capable of being expressed as Bessel or Bessel-Fourier series. A class of problems of considerable practical importance is that concerning local concentrations of stress. Such concentration of stress may be due directly to intense local loading; but more usually it is superposed on a simple uniform stress distribution by the disturbing effect of cracks, slits, holes, etc. The effect of a semi-infinite straight crack in an infinite plate has been investigated by Barnard[22] by compounding solutions due to equal and opposite forces or couples applied at adjacent points on opposite sides of the crack. Another straight crack problem has been treated by Gunder[23] who has investigated the distribution of shear stress in beams with slits parallel to the axis, the maximum shear stress in this case being stated to occur away from the end of the crack.

The case of a load applied to a narrow plate through a rivet hole, the load being either parallel or perpendicular to the length of the plate, has been treated by Knight[24]. He has determined the load distribution round the edge of the hole by considering the stress distribution set up at this radius in a solid plate by a load applied at the centre of the circle. This distribution is expressed as a finite Fourier series and a second distribution calculated to cancel the first round the "back" half of the edge of the hole, and to double it round the "front" half. This distribution results in stresses at the edges of the plate, which can be annulled by a further stress distribution. Using this process, which is rapidly convergent, the maximum stress is found to be from 3 to 5 times the mean stress in the plate. The stress concentration set up by an infinite row of equal and equally spaced circular holes in a plate, when the plate is loaded in tension or compression parallel or perpendicular to the line of the holes, has been considered by Howland[25]. He has defined harmonic and biharmonic functions having infinities at the points $x = \pm pb$, $y = 0$, p being an integer or zero and b a constant, the functions being analytic in the rest of the plane and periodic with period b. From a series of such functions, he has compounded a stress function which satisfies the conditions that the edges of the holes should be free from stress. The analysis leads to an infinite series of linear equations, and these equations are solved by a process of successive approximation, which is shown to be convergent provided the diameter of each hole

is not more than about half the spacing of the centres. When the stress is applied parallel to the line of holes the mutual shielding effect of the holes reduces the stress concentration effect in comparison with that due to a single hole, whilst, when the stress is applied perpendicular to the line of holes, the stress concentration effect is slightly increased. For holes of which the diameter is half the spacing the concentration factors are 2·16 and 3·24 as against 3 for a single hole.

Other problems related to cases of stress concentration have been treated by Hovgaard[26] and by Shepherd[27]. Hovgaard, dealing with welded joints, has considered the distribution of shear stress across the joint between a plate and a welded-on side bar due to the transference of tensile load from the plate to the side bar. He concludes that the shear stress at any point is proportional to $x/\sqrt{(L^2-x^2)}$ where $2L$ is the length of the joint and x is the distance of the point from the mid-section of the joint. Shepherd has treated the problem of an infinite sector loaded by concentrated forces normal to the faces of the sector at a finite distance below the apex. The problem is treated as one of plane stress and the solution is obtained by composition of exact solutions. Two cases are considered: in both, two forces are applied, one on each face at equal distances from the apex; in one case both forces are directed inwards, in the other one force is directed inwards and the other outwards.

The stresses set up in elastic bodies may not be due merely to the application of certain external forces; stresses may in some cases be set up by body forces due to acceleration or by differential thermal expansion. Considering thermal stresses, Biot[28] states that, if a solid cylinder is heated in such a way as to set up a steady state temperature distribution the same for all cross sections, the only stress set up is a tension normal to the cross section and proportional to the temperature at each point of the cross section. Biot states also that under certain conditions the same theorem applies to the case of a hollow cylinder.

The stresses set up in a rotating disc by the centrifugal forces constitute a problem of considerable practical interest and the distribution of stress in certain cases is well known. Sen[29] has recently obtained solutions in series for two further cases of discs whose thickness decreases radially either exponentially or parabolically. Problems of vibration also constitute a class in which the stresses are due entirely to body forces; but there is also an intermediate class of problem in which both body forces and applied forces play closely interrelated parts. This class of problem arises when a deformable body is acted upon by forces which cause appreciable acceleration of the whole or part of the body. Under the title of "the mechanics of deformable bodies" this class of problem has recently been considered by Weissenberg[30]: but the subject could equally well be termed "the deformation of movable bodies".

The mechanics of deformable bodies may be considered to include all problems in which the impressed forces, themselves functions of the deformation, cause not only deformation but also acceleration of portions of the structure. In certain cases, the impressed forces are all proportional to the same parameter and the chief interest lies in the determination of the minimum value of the parameter which causes the static configuration of the structure to become unstable. Duncan[31] has

recently published a proof for conservative systems of Minelli's method for determining this limit of stability. Minelli's method consists in the equation of the work done by the constant forces to the work done by the variable forces, the deformation being so chosen that the value of the parameter defining the scale of the variable forces should be a minimum. Duncan has shown how the method may be applied to determine the speed at which a certain type of instability may occur in an aircraft wing.

When the externally applied forces are zero or constant, the mechanics of deformable bodies reduces to the theory of vibration. Problems of this type, involving special problems of elasticity, have been investigated by Jagger[32], Lowan[33], Meyer zur Capellen[34] and Tomotika[35]. Jagger has treated the longitudinal vibration of an elastic bar by expressing the axial load distribution as a Fourier series; his method allows the effect of damping to be taken into account. Lowan has developed an operational method for the determination of the mode of oscillation of a simply supported beam under the action of forces varying both in space and time, whilst Meyer zur Capellen has considered the general case of free transverse oscillations of a stiff string under any end load and with either end either simply supported or clamped. The modes and frequencies of transverse vibration of a stretched string and of a supported beam are found as limiting cases of the general case; and by considering the case when a supported beam is subjected to compressive end load, the buckling load, i.e., the load at which the period of transverse vibration becomes infinite, is determined. Tomotika has considered the free transverse vibration of a square plate clamped at all four edges. The solution of the problem is obtained by the composition of exact solutions of the differential equation governing the motion, which themselves satisfy one boundary condition, in such a manner as to satisfy the other boundary condition at one frequency. The frequency of oscillation is then found by successive approximation to the roots of an infinite determinant, the process being very rapidly convergent.

It has already been mentioned that instability of any structure under load implies that the period of oscillation in that mode of deformation has become infinite. In practice it is of course important to know at what value of the applied loads such instability may be expected. Some few problems of this type have been solved by exact methods; but in general it is usually easier to assume a form for the mode of buckling and to find by application of the Rayleigh principle of least energy the load which will cause instability in this mode. The same type of considerations apply as in the application of the Rayleigh principle to problems of vibration; the value found for the critical load is always too high, but the error is usually quite small even when the true buckled form is very different from that assumed. Problems of the stability of thin rectangular plates under compressive stresses in the plane of the plate have been considered by Trefftz[36], Weinstein[37] and Leggett[38]. Both Trefftz and Weinstein have treated the case in which the plate is subjected to uniform compression at all four edges. They have obtained approximate solutions by the Rayleigh method and have shown how lower limits to the buckling load may be obtained by considering cases in which the edge conditions are less stringent (cf. Temple and

Bickley[39]). Trefftz has shown also that the approximate method leads to the same buckled form as that previously found by the exact method by Taylor[40]. Leggett, using the exact method, has considered the stability of a rectangular plate when loaded by compressive forces distributed over two opposite edges and varying linearly from zero at one corner. He has obtained solutions both for the case when the edges parallel to the load are simply supported and for the case when they are clamped, the loaded edges being simply supported in both cases. Beschkine[41] has stated some general conditions governing the effect of boundary conditions on the equilibrium of thin surfaces of any form. A problem related to that of buckling has been considered by Reissner[42]. The stress distribution set up by bending in the flanges of beams of T-section may be affected by variation in the displacement of the flange across its width. This variation causes the stress in the outer parts of the flange to be lower than the stress in the centre and reduces the "effective width" of the flange. Reissner concludes that continuation of the beam beyond the points of support tends to increase the effective clamping at the supports and also to increase the effective width of the flange at the weakest section.

The determination of the distribution of forces between the individual members of a framework under given systems of load forms an important practical branch of the theory of elasticity. If all the members of the framework may be considered pin-jointed at their ends and if the framework as a whole and in all its parts is simply stiff, the solution of the problem may be obtained very simply by straightforward methods; but, if the framework includes redundant members, or if it is desired to take into account the rigidity of the joints in resisting relative angular movements of the ends of the members, exact solutions cannot be obtained by straightforward step by step methods. A framework is said to be simply stiff when the number and disposition of its bars is just, and only just, sufficient uniquely to determine the position of each joint relative to the others. A redundant member is one which connects two joints which are already fixed in relative position. Thus in a plane the simplest simply stiff framework is that formed by three bars in the form of a triangle, and if a fourth joint is connected to two of the joints of this triangle by two (non-parallel) bars, the resulting framework is again simply stiff. In general each added joint requires the addition of two further bars, so that a necessary condition that a plane framework should be simply stiff is that the number of bars m should be related to the number of joints j by the relation $m=2j-3$. Similarly the simplest space framework that is simply stiff is that formed by six bars joining four joints in the form of a tetrahedron, and the addition of each new joint requires three further bars, so that one necessary condition that the framework should be simply stiff is that $m=3j-6$. However, these conditions are not sufficient to establish that a given structure is simply stiff, because it may include redundant members in one part and be a mechanism in another. Moreover, the structure may not be built up from the simplest possible framework. Roxbee Cox[43] has devised methods of synthesizing and analysing simply stiff frameworks. He has shown that a simply stiff plane framework may include irreducible units in which two or three bars meet at each joint, the latter being a hexagon braced by three diagonal bars, and that a space

framework may include irreducible units in which three, four or five bars meet at each joint, the latter case being an icosahedron consisting of twelve joints braced by 30 bars.

The presence of redundant bars in a framework, and in the present sense stiff joints may be considered as equivalent to redundant bars, may cause the framework to be self-stressed; but in addition the distribution of load between the various bars due to given external loads cannot be determined without considering the deformation of the framework. One method of solution is to remove sufficient bars to render the framework simply stiff and to represent the effect of the bars removed by unknown forces at the joints they should connect. By equating the relative movement of these joints to the deformation of the bars removed under the forces assumed to act in them, a number of simultaneous equations equal to the number of redundant bars is obtained. If the number of redundancies is large, this method becomes extremely laborious; but the more serious objection to it is that it becomes difficult to obtain a sufficiently accurate solution of the simultaneous equations. Accordingly alternative methods of treating such problems have been devised. Southwell and Owen[44] have devised a method for analysing the forces in a stiff-jointed radially-braced polygonal ring loaded at the joints by forces in its own plane. They show that, for certain types of distribution of load, exact and fairly simple solutions may be obtained, and they prove that any load distribution can be represented to any desired degree of accuracy by composition of appropriate load distributions of special type.

For the solution of general problems of this type, particularly applicable to the problem of stiff jointed frameworks, a new approximate method has recently been devised by Hardy Cross[45]. This method has been developed by Baker and Ockleston[46], Morris[47], Williams[48] and Southwell[49]. In this method, the process of successive approximation is associated with a definite sequence of physical operations which could in theory actually be performed on the actual structure. The process consists in subjecting the framework to a system of restraints chosen in such a manner as to leave a problem soluble by ordinary means, whilst at the same time leaving the major actions unrestrained. The solution of the simplified problem indicates *inter alia* the forces and moments on the artificial constraints. Any one of these constraints may then be considered to have been removed if to the framework forces and moments equal and opposite to those imposed on the constraint are applied. Readjustment of both the internal loads and the reactions on the other constraints due to these new applied forces can then be calculated and the whole process can be repeated until the residual restraints are negligible. In practice the method proves to be rapidly convergent.

When the redundancy of a framework consists in the rigidity of the joints, i.e. in the possibility of transmitting a bending moment from one bar to another, a convenient method of ascertaining the forces and moments consists in equating to zero the relative displacements or rotations, expressed as differential coefficients of the total strain energy in the framework, between fixed points. In applying this method it is usual to neglect the direct and shear strain energy in the members in

comparison with that due to bending. Bateman[50] has investigated the effect of including the direct strain energy; in the case of a typical portal frame the effect on the bending moments is shown to be less than 2 per cent, whilst in a more unfavourable case an error of 7½ per cent is found.

References

(1) Seth, B. R. *Philos. Trans.* A, **234**, 231–64 (1935).
(2) Phillips, H. B. *J. Math. Phys.* **13**, 421–5 (1935).
(3) Bateman, E. H. *Phil. Mag.* **19**, 1011–15 (1935).
(4) Seth, B. R. *Phil. Mag.* **20**, 632–40 (1935).
(5) Seth, B. R. *Proc. Camb. phil. Soc.* **30**, 392–403 (1933–4).
(6) Okubo, H. *Proc. phys.-math. Soc. Japan*, **16**, 430–5 (1934).
(7) Sunatani, C. *Technol. Rep. Tôhoku Univ.* **11**, 267–77 (1935).
(8) Sunatani, C. and Minagawa, S. *Technol. Rep. Tôhoku Univ.* **11**, 417–32 (1935).
(9) Deimel, R. F. *Proc. nat. Acad. Sci.*, Wash., **21**, 637–42 (1935).
(10) Schwalbe, W. L. *Z. angew. Math. Mech.* **15**, 138–43 (1935).
(11) Trefftz, E. *Z. angew. Math. Mech.* **15**, 220–5 (1935).
(12) Williams, D. *Rep. Memor. aero. Res. Comm.*, Lond., No. 1669 (1935).
(13) Föppl, L. *Z. angew. Math. Mech.* **15**, 37–40 (1935).
(14) Coker, E. G. and Filon, L. N. G. *Photo-Elasticity* (Cambridge University Press, 1932).
(15) Tuzi, Z. and Nisida, M. *Sci. Pap. Inst. phys. chem. Res.*, Tokyo, No. 566, 277–309 (1935).
(16) Gough, H. J., Cox, H. L. and Sopwith, D. G. *Proc. Instn mech. Engrs*, Lond., **128**, 253–360 (1934).
(17) Macgregor, C. W. *Bull. Univ. Pittsburgh*, **31**, 242–54 (1934).
(18) Habel, A. *Z. angew. Math. Mech.* **15**, 185–90 (1935).
(19) Bateman, E. H. *Phil. Mag.* **20**, 607–10 (1935).
(20) Vint, J. and Elgood, W. N. *Phil. Mag.* **19**, 1–21 (1935).
(21) Sen, B. *Phil. Mag.* **20**, 1158–63 (1935).
(22) Barnard, M. M. *Z. angew. Math. Mech.* **15**, 121–37 (1935).
(23) Gunder, D. F. *Physics*, **6**, 38–42 (1935).
(24) Knight, R. C. *Phil. Mag.* **19**, 518–40 (1935).
(25) Howland, R. C. J. *Proc. roy. Soc.* A, **148**, 471–91 (1935).
(26) Hovgaard, W. *J. Math. Phys.* **13**, 195–248 (1935).
(27) Shepherd, W. M. *Proc. roy. Soc.* A, **148**, 284–303 (1935).
(28) Biot, M. A. *Phil. Mag.* **19**, 540–9 (1935).
(29) Sen, B. *Phil. Mag.* **19**, 1121–5 (1935).
(30) Weissenberg, K. *Arch. Sci. phys. nat.* **17**, 44–106 and 130–71 (1935).
(31) Duncan, W. J. *Phil. Mag.* **20**, 789–93 (1935).
(32) Jagger, J. G. *Phil. Mag.* **20**, 997–1000 (1935).
(33) Lowan, A. N. *Phil. Mag.* **19**, 708–15 (1935).
(34) Meyer zur Capellen, W. *Z. angew. Math. Mech.* **15**, 243–5 (1935).
(35) Tomotika, S. *Rep. aero. Res. Inst. Tokyo*, No. 129, 301–28 (1935).
(36) Trefftz, E. *Z. angew. Math. Mech.* **15**, 339–44 (1935).
(37) Weinstein, A. *J. Lond. math. Soc.* **10**, 184–92 (1935).
(38) Leggett, D. M. A. *Proc. Camb. phil. Soc.* **31**, 368–81 (1935).
(39) Temple, G. and Bickley, W. G. *Rayleigh's Principle* (Oxford University Press, 1933).
(40) Taylor, G. I. *Z. angew. Math. Mech.* **13**, 147–52 (1933).
(41) Beschkine, L. *C.R. Acad. Sci.*, Paris, **201**, 935–7 (1935).
(42) Reissner, E. *Z. angew. Math. Mech.* **15**, 359–64 (1935).
(43) Roxbee Cox, H. *Proc. Lond. math. Soc.* **40**, 203–16 (1934–5).
(44) Southwell, R. V. and Owen, J. B. B. *Phil. Mag.* **20**, 1073–92 (1935).
(45) Cross, H. *Trans. Amer. Soc. chem. Engrs*, **96**, 1–10 (1932).

(46) BAKER, J. F. and OCKLESTON, A. J. *Rep. Memor. aero. Res. Comm.*, Lond., No. 1667 (1935).
(47) MORRIS, J. *Rep. Memor. aero. Res. Comm.*, Lond., No. 1672 (1935).
(48) WILLIAMS, D. *Rep. Memor. aero. Res. Comm.*, Lond., No. 1670 (1935).
(49) SOUTHWELL, R. V. *Proc. roy. Soc.* A, **151**, 56–95 (1935) and A, **153**, 41–76 (1935).
(50) BATEMAN, E. H. *Phil. Mag.* **19**, 59–66 (1935).

§ 3. VISCOSITY

By G. BARR

Kinematic viscosity: the stokes. It is always somewhat disconcerting to the student of general physics when specialists in one branch introduce new names for units, and it may be well to preface these notes by explaining that many of those who have to deal with the determination of viscosity for technical purposes have agreed [1] to report their results in terms of kinematic viscosity rather than in terms of *dynamic viscosity* (as the ordinary "viscosity" of the text-books is now distinguished). The symbol approved for dynamic viscosity being η, that for kinematic viscosity is commonly ν, so that $\nu = \eta/\rho$, if ρ is the density of the fluid under consideration. There are logical reasons for regarding the kinematic viscosity as the more important property, and in any case the great majority of viscosity determinations are actually, in the first place, determinations of kinematic viscosity. In lubrication work, oils are bought and sold on a grading that depends almost exclusively on their kinematic rather than on their dynamic viscosity, and it is only in the actual experimental comparison of frictional losses in an oiled bearing that the dynamic viscosity becomes the essential factor. In hydrodynamical equations the viscosity involved is normally the kinematic. The writer has suggested [2] that kinematic viscosity be distinguished by a specific name such as "lentor" so that there should be no confusion between two properties.

The c.g.s. unit of dynamic viscosity having received the name "poise" on the suggestion of Deeley and Parr (1913) it became desirable that the corresponding unit of kinematic viscosity should also have some convenient designation. Jakob [3] proposed that the unit be called the "stokes", in honour of the British physicist, and the name has now been adopted by standardizing institutions in most countries. Water at 20° C. has a kinematic viscosity a little greater than one centistokes.

Theory of the viscosity of liquids. The striking success that attended the development of the kinetic theory of gases in the last century led physicists to hope that an equally satisfactory and illuminating theory might be advanced to explain the properties of liquids and of solids. The present generation has seen many brilliant verifications of the lattice theory of the crystalline state, but speculations on the nature of liquids have been less fruitful. Observations of Brownian movements in colloidal suspensions have afforded visual evidence of random motion such as is postulated for the molecules of a gas, and dilute solutions have been shown to obey laws similar to the simple gas laws, but kinetic theories have so far proved to be inadequate to provide a reasonable picture of the constitution of liquids. In view of the difficulties that have been experienced in attempting to formulate any general relationships between the

viscosities of pure liquids at different temperatures and pressures or between the viscosities and compositions of solutions, it is perhaps surprising that the first success of a new suggestion for the basis of a theory of the liquid state should be the deduction of an approximate value of the viscosity of certain liquids from other known properties.

(*a*) Andrade[4] starts by considering that a liquid differs less from a solid than it does from a gas: arguments derived from the kinetic theory of gases, in which molecules are separated by distances that are, on the average, very great compared with their own dimensions, are hence less likely to be helpful than analogies between the forces operative between molecules in solids and liquids. The essential difference, he suggests, between the two denser states, "is not the magnitude of the intermolecular force under which the molecule vibrates, but the amplitude of the motion". He first assumes, therefore, that a molecule of a liquid vibrates, with a frequency (approximately?) equal to that occurring in the solid state, about a mean position that is slowly displaced instead of remaining as in a solid fixed in relation to other molecules. The postulate of equality of frequency is supported by the similarity between the specific heats in the two phases of monatomic elements near their melting-points: he gives calculations which demonstrate that the number of collisions between the molecules of liquid lead is of the order of $4{\cdot}4 \times 10^{12}$ per second, so that the equilibrium position of a molecule of a liquid can move but slowly. He makes one other assumption, viz. that when two molecules come into contact with one another they may unite for a time which, though very short compared with the period of vibration, is yet long enough to ensure that they shall have the same velocity of translation: near the freezing-point the probability of union occurring at every close approach is nearly 100 per cent. If the molecules are in layers that are moving with different velocities, the temporary union of molecules in different layers entails a communication of momentum from one layer to another. This mechanism of transfer of momentum thus differs radically from that assumed in Maxwell's theory of gaseous viscosity, which depends upon an actual interchange of molecules between the layers: as a first approximation, at any rate, Andrade regards the displacements of the equilibrium positions of molecules as contributing only a negligible proportion of the momentum transferred.

A simple calculation then leads to a value for the viscosity η in terms of the masses of the molecules, the average distance between their centres and the frequency of vibration. For a liquid element near its freezing-point he takes the frequency of vibration of the molecules as being given by the formula proposed by Lindemann for solids:

$$\text{frequency} = \frac{C}{V_A^{\frac{1}{3}}}\sqrt{\frac{T_m}{A}},$$

where C is a constant, A the atomic weight and V_A the volume of a gram-atom at the melting-point T_m. After inserting accepted values for C and for Avogadro's number, the formula for the viscosity becomes

$$\eta = 5{\cdot}1 \times 10^{-4}\,(AT_m)^{\frac{1}{2}}/V_A^{\frac{2}{3}} \qquad \text{.....(1).}$$

It may facilitate comparison with recorded values if the formula is altered to

$$\eta = 5{\cdot}1 \times 10^{-4}\, T_m^{\frac{1}{2}}\, \rho_m^{\frac{1}{3}}/A^{\frac{1}{2}} \qquad \ldots\ldots(1a),$$

where ρ_m is the density (g./ml.) at the melting-point.

Andrade finds unexpectedly close concordance between his formula and experimental values for liquid mercury, lead, tin and copper: for antimony and bismuth (which crystallize in a less closely packed structure) the discrepancies are still only 37 and 22 per cent. Even for liquid chlorine, bromine, iodine, oxygen and hydrogen the formula yields results that are of the correct order.

(*b*) In the second part of his paper, Andrade[5] discusses the application of his theory to the effects of temperature, pressure and chemical constitution on viscosity. The assumption of a Boltzmann distribution formula for the relation between absolute temperature and the number of molecules possessing the critical energy necessary for the transfer of momentum by temporary union leads to equation (7) on page 23 as a first approximation. By considering the influence of thermal expansion on the processes involved he arrives at the modified formula

$$\eta v^{\frac{1}{3}} = Ae^{b/vT} \qquad \ldots\ldots(2),$$

in which v is the specific volume at the absolute temperature T. For over ninety pure liquids listed the data are probably represented within the error of the experiments, but with water and two tertiary alcohols the discrepancies may exceed 20 per cent. The existence of associated molecules does not, in general, prejudice the fit: if the degree of association varies rapidly with temperature, the constant b tends to have a high value. Mercury shows an excellent fit and gives an exceptionally low b.

The occurrence of anomalies even among pure substances and the extent of the variation of b between different liquids suggest that the main defect of the theory is that in its present form it offers little basis for expressing quantitatively the effect of the "shape" of the molecules or the nature of their fields of force. The theory is restricted to pure liquids: formula (2) is presumably little more suited than (7) to express the variation of viscosity of oils with temperature.. In this connexion it may be appropriate to refer to the work of Staudinger and his collaborators without attempting, however, to discuss whether their results may suggest a direction in which Andrade's theory could be extended.

(*c*) In a general way it has been recognized that large molecules or extensive associations of small molecules are probably responsible for high viscosities. From a study of the viscosities of dilute solutions (in tetralin or in benzol) of substances of high molecular weight, Staudinger[6] has been able to establish some fairly definite relations between the length of the molecule and the effect of the solute on the viscosity of the solution. He first converts his viscosity determinations to specific viscosities η_{sp} which may be defined as the excess above unity of the ratio (viscosity of solution)/(viscosity of pure solvent). For dilute solutions of highly polymerized products containing separate long-chain molecules he finds that the molecular weight M, and hence the length of the chain, may be determined from the relation

$$\eta_{sp}/c = K_m M \qquad \ldots\ldots(3),$$

where K_m is a constant for each homologous series of polymers and c is the concentration in gram-equivalent-weights of the polymerizing radicle per litre (e.g. c is 1/14 of the concentration in g./litre of a hydrocarbon C_zH_{2z}). The solution must be so dilute that η_{sp}/c is constant: at higher concentrations, e.g. even in 1 per cent solutions of an extremely highly polymerized product, the sum of the "spheres of influence" of the separate molecules may be of the order of the volume of the solution, and the viscosity will rise very rapidly with concentration.

Although K_m may be different in different polymeric series, it is simply related to the number of carbon atoms in the polymerizing radicle that are directly concerned in the formation of the chain. In fact, K_m divided by this number is virtually constant ($=0{\cdot}85\times10^{-4}=K_\epsilon$) for many different series including paraffins $(CH_2)_z$ for which $K_m=K_\epsilon$, polyprenes $(C_5H_8)_z$ for which $K_m=4K_\epsilon$, polystyrols $(C_8H_8)_z$ for which $K_m=2K_\epsilon$. For approximate calculations in this connexion, all hydrocarbons may be regarded as being nearly $(CH_2)_n$: a hydrocarbon of molecular weight 10,000, for example, which would contain 714 carbon atoms if it were a paraffin, would have 770 carbon atoms in the molecule if it were a polystyrol. Then since a normal paraffin with n CH_2 groups has a specific viscosity, in 1·4 per cent solution in tetralin, of $K_\epsilon\times M\equiv K_\epsilon\times 14n=0{\cdot}85\times10^{-4}\times14n=1{\cdot}19\times10^{-3}\,n$, we may find the approximate chain length n of any hydrocarbon with filiform molecules from the specific viscosity of its 1·4 per cent solution, divided by $1{\cdot}19\times10^{-3}$. Such considerations have given estimates for the molecular weights of caoutchouc and balata that agree reasonably well with those calculated from osmotic pressure data. Products containing mixtures of polymers of very different molecular weights will of course give values of K_ϵ differing from the normal. Further, the specific viscosity of substances with less than 35 carbon atoms in the chain is found to depend on the nature of the solvent, particularly on its viscosity and on the extent of solvatation: ideally, the law

$$\text{Specific viscosity of 1·4 per cent solution} = kn \qquad \ldots\ldots(4),$$

where k is constant, holds only if the solute has a viscosity infinitely larger than that of the solvent, as would be the case for a solute with a long filiform molecule dissolved in a solvent with small spherical molecules. The viscosity is scarcely affected by the presence of small side-chains or by the chemical nature of the product: thus normal and iso-polymers, unsaturated hydrocarbons and their hydrochloride compounds, give the same specific viscosity in 1·4 per cent solution provided they have the same number of carbon atoms in the chain.

(*d*) It may be recalled that Poiseuille's law was deduced from experiments on the flow of water through tubes of small bore and that although one of the objects of the experiments was to find general regularities that should govern the flow of blood through the veins and capillaries of the body, Poiseuille's law does not hold for blood. Although physicists may measure the viscosity of all sorts of pure liquids, it is rather with impure liquids, semi-liquids and semi-solids, that we have to deal in every-day life. Of these a large proportion may be described as "non-Newtonian", i.e. the rate of shear is not proportional to the shearing stress, and they do not obey Poiseuille's or Stokes's law. During the last few years[7] a large number of

observations have been made on the behaviour of such materials, including suspensions and emulsions, and the "apparent viscosities" have been measured and shown to vary not only with rate of shear but also, in many cases, with thermal and mechanical pretreatment. In the simpler examples, a given rate of shear in the system corresponds with a shearing stress that remains constant when the experiment is repeated: the data obtained in such cases have been carefully analysed and suitable mathematical expressions have been deduced to correlate them, but no physical explanation has yet been advanced that is general enough and at the same time capable of being sufficiently detailed to serve as a working hypothesis for all the systems considered. This is hardly surprising in view of our lack of knowledge of the molecular mechanism of viscous effects in simple liquids: it may well be that no startling advance will be made until conceptions arising from the study of non-Newtonian liquids have been allowed to influence the general theory of viscosity of liquids.

The variation of viscosity with temperature. A vast amount of labour has been expended in attempts to find general formulae that will reproduce the variations of viscosity with temperature and pressure and, in the case of solutions, with concentration. Although a few of the formulae proposed have had some sort of theoretical basis, most of them have been frankly empirical, and those that have given reasonable concordance with experiment over a large range of viscosity have usually required three or more arbitrary constants. Thorpe and Rodger for example applied to their results a formula suggested by Slotte[8]

$$\eta = b/(a+t)^n \qquad \ldots\ldots(5),$$

in which t is the temperature in °C. and b, a and n are constants each of which has a large range of variation between different liquids. McLeod[8] and Batschinski[8] were able to represent the same results by formulae relating the viscosity to the specific volume v (the coefficients of expansion with temperature being known); Batschinski's formula is

$$\eta = b/(v-a) \qquad \ldots\ldots(6),$$

where b and a are constants, but McLeod provides for an exponent of the bracketed term that may differ from unity for associated liquids.

Arrhenius used an equation making $\log \eta$ proportional to the concentration of solute in solutions, but it is only in later years that formulae of an exponential type have become popular for relating viscosity to temperature. The simplest, viz.

$$\eta = Ae^{b/T} \qquad \ldots\ldots(7),$$

in which A and b are constants and T is the absolute temperature[9], has been put forward independently by several investigators and may be deduced theoretically from a number of conceptions of the nature of the effect of temperature on viscosity. It holds fairly well for many pure liquids but breaks down entirely for viscous oils[10]. For such materials as oils and bitumens, of high viscosity with large temperature-coefficient of viscosity, the greatest accuracy of interpolation between results at different temperatures appears to be offered by Vogel's[11] equation which may be written in the form

$$(a+\log \eta)\,(T+b) = c \qquad \ldots\ldots(8).$$

If $b=0$ this reduces to equation (7). Cragoe[12] has shown that an equation with only two arbitrary constants, reducing to (8) if a is made always equal to 3·3, is capable of representing most of the published data on the variation of viscosity of oils with temperature.

Two other empirical logarithmic formulae may be mentioned here in view of their successful application over very wide ranges of viscosity. The first,

$$\log \eta = a - b \log \theta \qquad \text{.....(9)},$$

where θ is the temperature °F. was adopted by Evans and Pickard[13] for tars, for which Lee and Mitchell[14] have confirmed its applicability: it has also been used quite often for oils[15]. Actually, like many other empirical formulae, this is a special form of Slotte's equation (5), from which it may be derived by putting $a = 32 \times 5/9$.

For substances of extremely high viscosity, the rate of variation with temperature is so rapid that a plot of $\log \eta$ against temperature, or against any simple function of temperature, at once suggests the desirability of further contracting the viscosity scale as the viscosity rises. Le Chatelier[16] found that the viscosities of glasses in their working and annealing ranges of temperature could be represented by the equation

$$\log \log \eta = a - b \log T \qquad \text{.....(10)},$$

if the viscosity be expressed in poises. Such an equation has the peculiar defect that a change of units alters the shape of its curve and it will obviously cease to be of use at very low viscosities. The importance of the latter objection may be minimized by adding a suitable constant to the value of the viscosity before taking the first logarithm, as suggested by Walther[17] and others. If the kinematic viscosities of lubricating oils are expressed in centistokes (100ν), it is found that the formula

$$\log \log (100\nu + 0{\cdot}8) = a - b \log T \qquad \text{.....(11)}$$

holds with an accuracy very little inferior to that of Vogel's three-constant equation so long as viscosities less than about 5 cs. are excluded. The American Society for Testing Materials[18], in accordance with a tentative specification, issues co-ordinate paper in which the scales are graduated to suit a similar formula for the purpose of interpolating kinematic viscosities of oils between values obtained at two different temperatures. Ubbelohde[19], who also advocates the use of a double logarithm chart, has found it to be of very great use in combining the results of tests on tars and bitumens over an enormous range of viscosity.

Viscometry. A list of the numerous forms of viscometer that have been devised since the publication of the writer's *Monograph of Viscometry* (1931) would probably be out of place here, but it may be well to draw attention to a recent study of the wall-correction in the falling-sphere viscometer. Ladenburg's correction to Stokes's equation has been assumed so long that Faxén's demonstration of its inaccuracy has not yet succeeded in displacing it from use, although the more accurate formula and a verification of it by Schmiedel were quoted in the above-mentioned monograph. Bacon[20] has just published a paper describing experiments which covered a wide range of Reynolds's numbers R and of diameters d, D of sphere and tube: these

confirmed the accuracy of the Faxén correction, so that if η_s is the viscosity calculated from Stokes's equation for free fall, the true viscosity is $\eta_s F$, where

$$F = 1 - 2{\cdot}104\, d/D + 2{\cdot}1\, (d/D)^3 - 0{\cdot}95\, (d/D)^5 \qquad \ldots\ldots(12).$$

The mathematical derivation assumes that R^2, $R\,(d/D)^2$ and $R^2\,D/d$ are negligible compared with unity.

References

(1) Cf. resolution of viscosity section, *Proc. World Petroleum Congr.* **2**, 543 (London, 1933).
(2) Barr. *Proc. World Petroleum Congr.* **2**, 508 (London, 1933).
(3) Jakob. *Z. tech. Phys.* **9**, 21 (1928).
(4) Andrade. *Phil. Mag.* **17**, 497 (1934).
(5) Andrade. *Phil. Mag.* **17**, 698 (1934).
(6) Staudinger, over 100 papers. See *Ber. dtsch. chem. Ges.* **65**, 267 (1932) and *Z. Elecktrochem.* **40**, 434 (1934).
(7) The *Journal of Rheology* (1930–2), now merged into *Physics*, contained many papers and a bibliography of the subject.
(8) References in Hatschek's *The Viscosity of Liquids* (London, 1928).
(9) Andrade. *Nature*, Lond., **125**, 580 (1930).
(10) Erk and Eck. *Phys. Z.* **37**, 113 (1936).
(11) Vogel. *Phys. Z.* **22**, 645 (1921).
(12) Cragoe. *Proc. World Petroleum Congr.* **2**, 529 (London, 1933).
(13) Evans and Pickard. *An Investigation into the Nature and Properties of Coal Tar.* (South Metropolitan Gas Co., London, 1931).
(14) Lee and Mitchell. *J. Soc. Chem. Ind.*, Lond., **54**, 407 T (1935).
(15) Herschel. *J. Industr. Engng Chem.* **14**, 715 (1922).
(16) Le Chatelier. *Ann. Phys.*, Paris, **3**, 5 (1925).
(17) Walther. *Maschinenbau*, **10**, 671 (1931); reference from Erk, *loc. cit.*
(18) *A.S.T.M. tent. Stand.* p. 775 (1935).
(19) Ubbelohde. *Zur Viskosimetrie* (Berlin, 1935).
(20) Bacon. *J. Franklin Inst.* **221**, 251 (1936).

FLUID MOTION

By A. FAGE, A.R.C.Sc., F.R.Ae.S.

§ 1. INTRODUCTION

The subject of fluid motion covers a wide field, from the problems of slow motion, controlled by viscous forces, to those of very high speed, controlled by inertia and compressibility forces. Intermediate between these extremes are the motions controlled, in the absence of compressibility, by viscous and inertia forces, to the same order of magnitude. Mathematical solutions for certain slow motion problems can be obtained from the equations of motion, simplified by the neglect of the inertia terms, and also for a few special problems of compressible inviscid fluid motion. Exact mathematical solutions for problems of intermediate speed have not been obtained, but a simplification suggested by Prandtl, coupled with experimental experience, often leads to reliable results.

Considerable progress has been made, within recent years, in the study of fluid flow phenomena. A survey covering the entire field of progress is impossible within the compass of a short article, and for this reason attention will be confined to some motions of the intermediate kind, which are controlled by both viscous and inertia forces, and to turbulent motion. Personal predilection has influenced, to some extent, the choice made, but nevertheless it is an important one, for these motions are the commonest of those experienced in nature, and they have formed the subject of many theoretical and experimental investigations. Although the field of survey has been purposely narrowed by the exclusion of important borderland subjects—slow motions, flow of compressible fluids, meteorology, tides, etc.—the ground is far too extensive to permit a detailed exploration, and there appears no alternative but to outline some of the outstanding features, and to hope that these features will convey some idea of the progress made. Further an endeavour has been made to present the survey in a form which can be followed not only by the expert, but also by the physicist who has no specialized knowledge of fluid motion. It is perhaps hardly necessary to add that the bibliography given is far from complete, and that references to many important papers are given in the papers cited.

§ 2. FLOW PATTERN

When a body is totally immersed in a steady fluid stream the resistance depends solely, in the absence of compressibility effects, on the non-dimensional Reynolds' number Ul/ν, where U is the speed, ν is the kinematic viscosity obtained by dividing the coefficient of viscosity μ by the density ρ, and l is a linear dimension, usually the downstream length of the body. The resistance of a body arises from the normal pressures and the tangential stresses on its surface. A streamline body

is one whose attitude to the stream and whose shape are such that the greater part of the resistance arises from the tangential stresses.

At very low Reynolds' number, the flow past a streamline body is everywhere laminar, and the influence of viscosity is spread over the entire field. With an increase in Reynolds' number, the region over which the direct effects of viscosity extend becomes smaller, and eventually it is confined to a thin layer of fluid adjacent to the surface of the body. This thin layer is called the boundary layer. The motion of the fluid within this layer is controlled partly by dynamic forces associated with the kinetic energy and the pressure of the fluid, and partly by frictional forces associated with the viscosity. The frictional forces preponderate very close to the surface, and on moving outwards from the surface they become less important, and eventually negligible at and beyond the outer limit of the layer. Any work done against frictional forces is at the expense of the total pressure energy, and the boundary layer can be taken to be the region adjacent to the surface where there is a loss of total pressure head.

On the nose of a body, the flow in the boundary layer is laminar, but eventually this flow breaks down and becomes turbulent. The position of the transition region depends on the Reynolds' number, on the turbulence in the general stream, and on surface roughness; and an increase in each causes an earlier transition. Energy can be easily communicated to a turbulent layer from the faster moving fluid just outside, and on a good form the turbulent layer adheres to the surface right up to the tail. On a short bluff body, energy cannot be absorbed by the layer fast enough to prevent the inner part being brought to rest, and the layer separates from the surface in the form of a free vortex sheet, which breaks up into eddies and forms a turbulent wake within which energy is rapidly dissipated.

In general, the flow in a boundary passes through three régimes—laminar, transitional, and turbulent. The distinction between laminar and turbulent flow in a boundary layer is not based on the presence or absence of velocity fluctuations, but on the fact that in laminar flow the transfer of momentum is by molecular action, whereas in turbulent flow the transfer is by small masses of fluid (molar action). There is also a marked difference in the velocity distributions, that for a laminar boundary layer resembling a parabolic form, whilst that for a turbulent layer approximates to a logarithmic form, except near the surface. Oscillograph records[1,2] show that comparatively slow velocity fluctuations occur in a laminar boundary layer, whereas the fluctuations in a turbulent layer are much more rapid. In the transition region there is a spasmodic change from laminar to turbulent flow, occurring at infrequent intervals near the beginning of the transition region, and more and more frequently towards the end of this region. The phenomenon resembles that observed by Osborne Reynolds in the flow of water through a pipe. Experiments[2] have shown that the resistance changes of a hot wire in a boundary layer, due to velocity fluctuations, can be made audible in a telephone receiver by suitable electrical apparatus, and that the character of the noise emitted changes as the wire is traversed through the transition region from laminar to turbulent flow.

§ 3. INVISCID FLOW

The intractable character of the equations of viscous flow compelled mathematicians of the last century to consider the simpler problems of a perfect fluid—the equations obtained when the viscosity terms are neglected are more amenable to treatment—and important theorems on the nature of inviscid flow possessing circulation and vorticity (the vorticity is supposed to be concentrated on sheets or along lines) were enunciated. These general theorems have proved invaluable in the development of modern theory, and from them specialized mathematical relations, capable of application to the problems of a real fluid, especially those experienced in aeronautics, have been obtained. A long cylindrical body placed in a perfect fluid with its axis normal to the stream experiences a lift when the flow is such that there is a circulation around the contour, expressed mathematically by $\int_{\text{contour}} v_s ds$, where v_s is the velocity component in the direction of the tangent ds of the contour. The relation between the lift per unit span, L, and the circulation, κ, is $L = \kappa \rho V$, where V is the velocity of the undisturbed stream. Kutta[3] in 1902, and Joukowski[4] in 1905, independently showed that the strength of the circulation (and so the lift) could be calculated, when the conformal transformation by means of which the section is derived from a circle is known, on the assumption that the streams above and below the body unite smoothly at the trailing edge. The relation of Kutta and Joukowski has valuable applications in aeronautical problems, particularly in the design of wing forms[5]. An experimental verification of this relation has been obtained by Bryant and Williams[6], and later by Tanner[7]. In the first investigation, a thorough exploration of the field of flow for two-dimensional motion past an aerofoil was made, with hot-wire and pressure-tube velocity meters, and the circulation was determined from these measurements. The lift was deduced from the measured distribution of the normal pressure on the surface. It was found that the circulation around the aerofoil was constant, within the limits of experimental error, and independent of the contour chosen, provided that the contour did not at any part approach too near the aerofoil, and also that the trailing wake was cut approximately at right angles to its core.

The pressure distribution on an aerofoil of infinite span and uniform section is the same at all sections. When the span is finite, and the aerofoil experiences a lift, the pressure distribution gradually changes from the mid-section towards the tips, and this change is associated with a circulatory flow at the tips from the high-pressure region below to the low-pressure region above the aerofoil. Surfaces of discontinuity are formed immediately behind the aerofoil where the lateral outflowing stream below the aerofoil makes contact with the lateral inflowing stream above the aerofoil, and ultimately these surfaces roll up into trailing vortices. These vortices are coupled to each other by the circulation around the aerofoil. At any point in the field of flow, there is an induced velocity due to this vortex system, and the changes of lift and resistance along the span, arising from the induced velocity, can be determined, when the forces for infinite span are known. These fundamental ideas, first stated independently by Lanchester[8–10] and Prandtl[11–13],

form the basis of modern aerofoil theory. Many elegant mathematical applications of this theory have been published, especially by Prandtl and his school, and the practical utility of the theory is now firmly established[5, 14]. The distribution of vorticity in the wake behind a finite aerofoil has been determined by Fage and Simmons[15] from velocity measurements taken with hot-wire and pressure-tube velocity meters. This investigation provides an experimental verification of the theoretical relation that the total strength of the vorticity leaving a semi-span of an aerofoil, as obtained by integration over a transverse plane close behind the aerofoil, is equal to the circulation around the median section; and it also shows that the distribution of vorticity behind the aerofoil is closely connected with the distribution of lift along the span. Photographs of the circulatory motion of air at aerofoil tips have been obtained by Caldwell and Fales[16], and other experimenters[17] have located, by direct measurement, regions of vorticity in the wake of an aerofoil, principally behind the tips.

§ 4. FLOW VERY NEAR A SURFACE

The frictional stresses on the surface of a body are intimately related to the flow in the boundary layer. When the flow in the layer is laminar, the frictional intensity rises to a maximum value at or near the nose of the body, and falls with an approach to the point at which transition begins. A rise occurs over the transition region, followed by a fall in the region beyond, where turbulence is fully established. The intensity of surface friction, f, can be directly determined from the well-known relation $f=\mu\,(\partial U/\partial y)_0$, where μ is the coefficient of viscosity and $(\partial U/\partial y)_0$ is the velocity gradient at and normal to the surface. In general, a boundary layer is very thin—the thickness at transition on a symmetrical aerofoil having a chord of 40 in. is about 0·1 in. at a wind speed of 80 ft./sec.—and the velocity changes rapidly near the surface, so that a reliable prediction of $(\partial U/\partial y)_0$ from velocity measurements can only be obtained when they are taken very close to the surface. Small Pitot tubes and hot-wire anemometers are commonly used to measure the velocity distribution in a boundary layer, but these instruments are not suitable for measurement very close to a surface; because a Pitot tube of the ordinary type does not allow a sufficiently close approach to the surface, even when the diameter of the tube is small, and heat is conducted across the thin air layer between a hot wire and the surface, and the correction for this loss is difficult to determine accurately. Velocity measurements very close to a surface can however be made with a special form of surface tube, designed by Stanton[18]. The design of this tube is such that the inner wall of the tube is formed by the surface itself, and that the width of the opening is varied by moving the outer wall, the smallest width being about 0·001 in. Owing to the extremely small width of the opening, the speed deduced from the measured pressure is not the same as that at the geometrical centre of the mouth of the tube, and calibration in a known laminar field is made to determine the position of the effective centre corresponding to this speed. Stanton's experiments with these tubes mounted on the inner wall of a pipe lead to the conclusion that in turbulent motion there exists at the surface a thin layer of fluid within

which the flow is laminar in character, and the velocity at the surface is zero. Expressed in symbols, this boundary condition is $\mu \operatorname{Lt}_{y\to 0}(dU/dy)=f$. Surface tubes have also been used[19] to determine the intensity of friction on a symmetrical aerofoil and a circular cylinder.

In turbulent flow, the scale of turbulence near a solid surface is small, and the method used to observe the flow must not interfere with the internal motions of the fluid. The method of examining fluid motion with an ultramicroscope used at the National Physical Laboratory satisfies this condition, since measuring instruments are not introduced into the fluid. The first experiment[20] undertaken was the examination of turbulent flow in a square pipe of 0·89 in. side. A beam of intense bright light passing through a window in one side of the pipe illuminated particles contained in water flowing through the pipe, and the motions of these particles as they travelled downstream were viewed with an ultramicroscope through an adjacent window. Particles passing through any fixed point move in different directions, when the motion is turbulent, and the short traces of these paths seen in the microscope appear as bright rectilinear streaks inclined at various angles to the mean direction of flow. The maximum inclinations of these streaks can be measured, and the mean speed and the maximum speed fluctuations in the axial direction can also be measured by a technique based on the fact that illuminated particles are brought to apparent rest by motion of the objective of the microscope at the appropriate speed. It was found that the angular deviation, from the mean direction of flow, of particles in an axial plane normal to the wall increased slowly with the distance from the axis and reached a maximum value at about 0·05 in. from the wall. The deviation then fell to zero at the wall, so that particles seen when looking along and very close to the wall appeared to be moving in laminae parallel to the wall. The maximum deviation in planes parallel to the wall increased progressively as the wall was approached. A view of a surface layer of thickness 0·001 in. under high magnification showed that the faster-moving particles in this layer were moving in sinuous paths, and careful observation showed that the slower ones, some of which were estimated to be not more than $2{\cdot}5\times10^{-5}$ in. from the surface, crossed and recrossed several times a hairline in the eyepiece of the microscope. Occasionally, when a group of the faster particles moved laterally sideways, the slower particles would also shift laterally. It appears that when the general motion is turbulent the flow in a very thin surface layer is not quite regular. Nevertheless, the friction on the surface can be determined from the relation $f=\mu\,(dU/dy)_{y=0}$ for laminar flow. A recent examination of turbulent flow in a circular pipe[21] has shown that velocity measurements taken with the ultramicroscope technique can be linked up satisfactorily with those taken, very near the wall, by Stanton with surface tubes; and that the thickness of the thin layer, within which the velocity increases linearly with the distance from the surface, appears to be about $2{\cdot}5\nu/U_*$, where U_* is the frictional velocity given by $(f/\rho)^{\frac{1}{2}}$. The value of U_* for water flowing through a 1 in. circular pipe at a mean speed of 10 ft./sec. is about 0·5 ft./sec., and the corresponding thickness of this thin layer at normal temperature is about 7×10^{-4} in.

§ 5. SURFACE ROUGHNESS

Surface roughness affects the resistance to flow, and the influence of a systematic change in wall roughness on the flow in a circular pipe has been investigated recently by Nikuradse[22]. The roughness was obtained by the adhesion to the walls of sand grains, graded so as to obtain uniformity of size and shape, and the roughness was specified by the ratio of the mean height of the excrescences, k, to the radius of the pipe, r. The range of k/r covered was from 0·002 to 0·067. This scale of roughness had practically no effect when the flow was laminar, but at a high Reynolds' number when the flow was turbulent the resistance was appreciably increased. The resistance curves for the rough pipes departed in order of roughness from that for a smooth pipe, as the Reynolds' number was increased; and at a low value of (k/r) the resistance was practically the same as that for the smooth pipe, over a considerable range of Reynolds' number. The explanation offered by Prandtl is that excrescences do not affect resistance when they lie within the thin laminar sub-layer at the wall; but when they begin to project beyond this layer, small vortices are formed, and additional energy is lost. At a sufficiently high Reynolds' number, the resistance coefficient becomes constant, because the laminar sub-layer is so thin that all the excrescences project beyond it, and the energy loss associated with the vortices becomes constant. Prandtl and Schlichting[23] have deduced from Nikuradse's pipe results the resistance laws for a plate having the same form of surface roughness. It is shown that roughness does not affect the resistance until the average height of the excrescences exceeds $3\nu/U_*$, which is, as we have just seen, roughly equal to the thickness of the thin laminar sub-layer at the surface.

§ 6. THEORY OF LAMINAR BOUNDARY LAYER

The rate of change of shearing stress in a laminar boundary layer is large, and the viscous terms in the equations of motion are comparable with the inertia terms, even when the viscosity is small. The general equations of motion can be greatly simplified, in the manner first suggested by Prandtl[24] in 1904, and from these simplified equations the flow near a solid boundary, or in a laminar wake, can be closely predicted. The boundary layer equations for steady two-dimensional flow past a surface are

$$u\frac{\partial u}{\partial x}+v\frac{\partial u}{\partial y}=-\frac{1}{\rho}\frac{\partial p}{\partial x}+\nu\frac{\partial^2 u}{\partial y^2},$$

$$0=-\frac{1}{\rho}\frac{\partial p}{\partial y},$$

and the equation of continuity is

$$\frac{\partial u}{\partial x}+\frac{\partial v}{\partial y}=0,$$

where y is the distance measured normal to the surface, x is the distance measured along curves orthogonal to the normals, u and v are the fluid velocity components

in the directions x and y increasing, and p, ρ, and ν are the pressure, density, and kinematic viscosity of the fluid. The pressure gradient is independent of y, so that the term $\partial p/\partial x$ can be replaced by the pressure gradient at the edge of the boundary layer. A solution of the boundary layer equations can be obtained when $\partial p/\partial x$ is known, either from experiment or calculation. The earliest and one of the best methods of solution, due to Blasius[25] and Hiemenz[26], was obtained by expressing the velocity in the main stream at the edge of the boundary layer as a power series in x. Another method, primarily concerned with the determination of skin friction, and due to Kármán[27] and Pohlhausen[28], assumes that the velocity distribution can be expressed as a polynomial in y. Most of the methods tend to break down near the region where a boundary layer separates from the surface, and a modification of Pohlhausen's method, suggested by Dryden[29], can be used to obtain a closer approximation to the flow in this region. The methods of boundary layer analysis for steady two-dimensional flow are critically analysed in a paper by Howarth[30]. Recently, Piercy and Preston[31] have obtained a solution for the problem of a flat plate by an iterative method involving the use of gamma functions. The boundary layer equations for a body of revolution were first discussed by Boltze[32], and a solution of these equations for a sphere has been obtained by Tomotika[33].

An interesting application of the boundary layer equations has been made by Howarth[34] to predict the lift on an elliptic cylinder, immersed with its major axis inclined to a steady stream. On the assumption that the whole of the vorticity is confined to the boundary layer and wake, the velocities, and therefore the pressures, at the points of separation on the upper and lower surfaces are the same. The lift is determined from that particular circulation in the inviscid flow outside the boundary layer which makes the pressures at the points of separation on the upper and lower surfaces, determined from boundary layer theory, the same.

§ 7. STABILITY OF LAMINAR MOTION

Stability of laminar motion has formed the subject of many mathematical investigations. The method of attack is to examine, from the equations of motion, the history of a small disturbance superposed on the general motion. The motion is judged to be stable if every disturbance, whatever its type, tends ultimately to vanish, and neutral or unstable, if the disturbance persists or increases. The initial steady motion must be known, and for this reason attention has been restricted to those few cases in which solutions of the equations of motion are known. With any given form of small disturbance the criterion which distinguishes between an initial increase or an initial decrease in energy of the disturbed motion gives a definite number for the Reynolds' number of the motion. The critical number obtained depends however on the form assumed for the disturbance. Early investigations on this subject were made by Helmholtz[35], Kelvin[36], Rayleigh[37], Osborne Reynolds[38], and others. Later, Orr[39] found, by an extension of the method of Osborne Reynolds, the form of disturbance which gives, for flow between parallel

planes, the Reynolds' number below which all small disturbances initially decrease. The theoretical numbers obtained are well below the observed lower criterion, and it appears that when the Reynolds' number of the motion is between Orr's number and the observed lower criterion, disturbances can be imposed which initially increase but afterwards die away. A discussion of disturbances of pure laminar flow which increase very greatly initially and subsequently die away is given in Orr's paper. A motion which is unstable for small disturbances cannot remain steady for speeds higher than that at which instability sets in, and a motion which is stable for small disturbances may become turbulent for large disturbances. The flow between rotating cylinders, investigated by Taylor[40], is the only case in which instability has been proved and the predictions experimentally verified. A definite mode of disturbance appears when a calculable Reynolds' number of the flow is just exceeded, and the instability takes the form of alternate ring-shaped vortices, symmetrical about the common axis of the cylinders, and spaced a definite distance apart. It appears from this investigation that when the mean flow is such that only one mode of disturbance is just unstable, all the others being stable, this mode immediately makes its appearance. A critical review of theoretical investigations on the problem of stability of fluid motion, with a bibliography of the more important investigations, is given in a paper by Southwell and Chitty[41]. Recent experimental investigations on the origin of turbulence in pipes have been made by Naumann[42], Nikuradse[43], and Schiller[44].

§ 8. WIND TUNNEL TURBULENCE

Turbulence in the artificial stream of a wind tunnel can have a marked influence on the character of the flow past a body, particularly on the resistance of a streamline body and on the maximum lift of an aerofoil; and the interpretation of wind tunnel results is often very difficult, when they depend on the two variables, the Reynolds' number of the body and the turbulence in the stream. At present, the scale and magnitude of turbulence in a wind tunnel stream cannot be specified with exactness. A rough measure of the magnitude of turbulence is given by the ratio of the root-mean-square of the longitudinal speed fluctuations from the mean speed, $\overline{(u^2)}^{\frac{1}{2}}$, to the mean speed U at the point.

The value of $\overline{(u^2)}^{\frac{1}{2}}$ can be measured with a special form of hot-wire anemometer, using a wire of small diameter (0·0001 to 0·0005 in.), an amplifier, an electrical network to compensate for the thermal lag of the wire, and an alternating current milliameter[45–47]. The critical Reynolds' number of a sphere is also used as a turbulence number for the free stream. There is a critical range of Reynolds' number U_0D/ν over which the resistance coefficient C_D, where C_D is given by resistance $\div(\frac{1}{2}\rho U_0^2.\pi D^2/4)$, falls from about 0·5 to about 0·1. The critical number for the sphere is taken to be the value of U_0D/ν at which $C_D=0{\cdot}3$, and this number depends on the turbulence in the free stream. The critical sphere number depends not only on the scale and magnitude of the velocity fluctuations, but also on the diameter of the sphere. Taylor[48] has shown that a unique relation exists between

measurements of $(U_0D/\nu)_{\text{crit.}}$ and $\frac{(\overline{u^2})^{\frac{1}{2}}}{U_0}\left(\frac{D}{M}\right)^{\frac{1}{5}}$ taken in streams behind square-meshed grids, of mesh M, when the turbulence is isotropic. A third method of measuring turbulence is due to Schaubauer[49] and also to Simmons and Brown[50], and depends on the fact that the temperature distribution in the wake behind a wire of small diameter is a function of the turbulence in the general stream. The wire acts as a line source of heat, and at some distance downstream the thermal wake widens at a rate dependent on the turbulence, and the cross-stream distance between points where the temperature elevation is one-half the maximum elevation gives a suitable measure of the turbulence. Schaubauer shows that the correlation of this distance with values of $(\overline{u^2})^{\frac{1}{2}}/U_0$ is very good. The apparatus required is limited to a thermocouple and a sensitive galvanometer.

The above methods of turbulence specification do not, in general, allow the flow phenomena of different bodies to be closely correlated with the turbulence in the stream. Thus the maximum lift measured on certain aerofoils in the compressed-air tunnel at the National Physical Laboratory agree with observations taken, at the same Reynolds' number in flight, on aeroplane wings having the same shape, but the sphere critical numbers, 194,000 for the tunnel and about 365,000 for the free air, are different. Tests[51] in the compressed-air tunnel also show that different aerofoils react differently, as regards maximum lift, to changes in the turbulence of the tunnel stream. The action of turbulence on flow phenomena is very complex, and a satisfactory method of measuring turbulence will not be possible until we understand more completely the mechanism by which turbulence in the free stream affects the transition from laminar to turbulent flow in the boundary layer of a body.

§ 9. THEORIES OF TURBULENCE

Mathematical treatments of turbulent flow are based on a conception due to Osborne Reynolds[52] that the motion can be separated into a mean flow whose components are U, V, W, and a superposed turbulent flow whose components are u, v, and w, the mean values of which are zero. These means may be taken with regard to time at a fixed point, or with regard to one of the components at a given instant of time. In all cases of steady mean flow the means are taken at a fixed point, over a long period of time. When the mean flow is variable, it is necessary to assume that the fluctuations in u, v, and w are so rapid, that a significant mean value can be taken in an interval which is so short that the change in U during that interval can be neglected. The equations of the mean flow are the same as the ordinary equations of motion provided the Reynolds' apparent stress components $-\rho\overline{u^2}$, $-\rho\overline{v^2}$, $-\rho\overline{w^2}$, $-\rho\overline{uv}$, $-\rho\overline{uw}$, $-\rho\overline{vw}$, due to the momentum carried across the surface by the velocity fluctuations, are added to the mean values of the stresses p_{xx}, p_{yy}, p_{zz}, p_{yz}, p_{zx}, p_{xy}, due to viscous action. Three of the Reynolds' stresses depend only on the magnitude of one component of velocity, but the three components of shear stress depend on the magnitudes of two component velocities and on the correlation between them. The coefficient of correlation between u and v is taken, as in statistics, as $R_{uv}=\overline{uv}/(\overline{u^2})^{\frac{1}{2}}(\overline{v^2})^{\frac{1}{2}}$. When the turbulence is fully developed

the three components of the turbulent motion tend to become equal to one another, except near a solid boundary.

In statistical theories of turbulence an endeavour is made to represent the turbulent field in terms of the mean values and frequency distribution of quantities connected with the motion. Burgers[53] attempted to apply to turbulence the statistical methods developed in connexion with the kinetic theory of gases. This theory formed the subject of a separate Progress Report by Trubridge[54] and need not be considered here, except to mention that further development appears desirable, because the equations of motion are not taken into account. Other current theories are the momentum transport theory (Prandtl)[55–57], the vorticity transport theory (Taylor)[58, 59], and the statistical theory of turbulence (Taylor)[48]. In the first two theories a mixture length—somewhat analogous to the mean free path of a molecule in a gas—is used to relate the turbulent motion to the mean flow. Small masses of fluid are supposed to transport momentum in the one case and vorticity in the other, in the same way as heat and momentum are transported in a non-turbulent flow by molecular agitation. The small masses are supposed to leave the layer in which the property transported is that of the mean flow in the neighbourhood, and to travel a distance l in a direction transverse to that of the mean flow. At the point of arrival the masses are supposed to mix with the surrounding fluid so that the property transferred becomes identical with the average property of the fluid in that region. In the simple case of the transport of momentum in a turbulent field of flow for which the mean velocity U is parallel to the axis of x the fluid mass arrives with a velocity difference $-l\dfrac{dU}{dy}$, when l is small, and we have shearing stress $\tau = -\rho\overline{uv} = \rho\overline{vl}\left(\dfrac{dU}{dy}\right)$, if the pressure gradients in the fluid which accompany the transfer are ignored. Prandtl then takes v proportional to u, and writes $\overline{lv} = l^2\left|\dfrac{dU}{dy}\right|$, where the vertical bars indicate that the absolute value is to be taken, since $\overline{lv}$ must be positive. The rate at which momentum is communicated to unit volume of the fluid is given by

$$M = \frac{d}{dy}\left[\rho\overline{vl}\,\frac{dU}{dy}\right] = \frac{d}{dy}\left\{\rho l^2\left|\frac{dU}{dy}\right|\left(\frac{dU}{dy}\right)\right\}.$$

Both the coefficient of correlation between u and v, and a proportionality factor for the mixing length, are assumed to be included in l^2. It will be seen that $\rho\overline{vl}$ is virtually a coefficient of viscosity. The value of l can be calculated from experimental values of τ and U, and relations for l can be obtained from a correlation of theoretical distributions of mean velocity with those determined experimentally. A generalized version of the momentum transport theory applicable to cases where neither the mean nor the turbulent motions are confined to two dimensions has been given by Prandtl. When u and v are absolutely correlated and $\overline{u^2} = \overline{v^2}$, $|\overline{uv}| = (\overline{u^2})^{\frac{1}{2}}(\overline{v^2})^{\frac{1}{2}} = \overline{u^2}$, and we have $\tau = \rho l'(\overline{u^2})^{\frac{1}{2}}(dU/dy)$. The value of l' can be evaluated when τ and $(\overline{u^2})^{\frac{1}{2}}$ and dU/dy are known.

Prandtl has suggested that near a plane wall $l = By$, where y is the distance from the wall and B is a constant. Another hypothesis, due to Kármán[60, 61], is that l depends not directly on the distance from the wall but on the distribution of mean velocity, and the simplest relation for l on this hypothesis is

$$l = K \frac{dU}{dy} \Big/ \frac{d^2U}{dy^2},$$

where K is constant close to the wall. Prandtl's hypothesis leads to

$$\tau = B^2 \rho y^2 (dU/dy)^2$$

and the solution $$U = \frac{(\tau/\rho)^{\frac{1}{2}}}{B} \log y + \text{const.}$$

Kármán's hypothesis leads to

$$\tau = K^2 \rho \frac{(dU/dy)^4}{(d^2U/dy^2)^2},$$

and $$U = \frac{1}{K} \left(\frac{\tau}{\rho}\right)^{\frac{1}{2}} \log y + \text{const.}$$

These solutions are identical if $B = K$.

Taylor's vorticity theory somewhat resembles Prandtl's momentum theory, but differs from it in the very significant feature, that each small mass of fluid is supposed to retain the vorticity, but not the momentum, of the region from which it departed. The reason for concentrating attention on the transference of vorticity rather than of momentum is that when the motion is confined to two dimensions the local pressure differences do not affect the vorticity of a fluid element, but they certainly affect the momentum of individual masses of the fluid. The difference between the two theories can be illustrated by consideration of two-dimensional mean flow (plane XOY) parallel to OX. According to Prandtl's theory, the tangential stress is given, as we have already seen, by $\tau = \overline{\rho l v}\,(dU/dy)$. The rate at which momentum is communicated to unit area of a layer of thickness δy is

$$\rho \frac{d}{dy} \left[\overline{lv} \frac{dU}{dy}\right] \delta y,$$

and if the flow is due to a uniform pressure gradient $d\bar{p}/dx$, the equation for U is

$$\frac{d}{dy} \left\{\overline{lv} \frac{dU}{dy}\right\} = \frac{1}{\rho} \frac{d\bar{p}}{dx}.$$

When vorticity rather than momentum is transferred, the equation of motion, neglecting viscosity and on the assumption that the eddying motion is on the average uniform in the direction of x, leads to the relation

$$-\frac{1}{\rho} \frac{d\bar{p}}{dx} = \overline{2v\zeta},$$

where $$\zeta = \left(\frac{\partial v}{\partial x} - \frac{\partial u}{\partial y}\right).$$

Now the vorticity of the mean motion at any layer is given by $\frac{1}{2}(dU/dy)$, and on the hypothesis that vorticity is transferred without change, a small mass of fluid which has traversed a small distance l has a vorticity greater by an amount

$$-l\frac{d}{dy}\left(\frac{1}{2}\frac{dU}{dy}\right)$$

than that of the layer to which it has penetrated. Hence

$$2\overline{v\zeta} = -\overline{vl}\frac{d^2U}{dy^2}$$

and

$$\frac{1}{\rho}\frac{d\bar{p}}{dx} = \overline{vl}\frac{d^2U}{dy^2}.$$

The rate at which momentum is communicated to unit volume is given by

$$M = \rho\overline{vl}\frac{d^2U}{dy^2}.$$

This relation somewhat resembles that for Prandtl's theory, and the two relations are identical in the particular case when $\overline{vl}$ does not vary with y.

The momentum transport theory is identical with the vorticity transport theory for one type of turbulence only, namely when the turbulence is confined to the plane normal to the direction of mean flow, that is, when lines of particles parallel to the direction of the mean flow remain parallel to that direction. When the turbulent motion as well as the mean motion is confined to two dimensions, the vorticity transport theory yields results which are quite different from those predicted by the momentum transport theory. On the momentum transport theory the distributions of temperature and velocity in the wake at some distance behind a heated cylindrical obstacle are identical, while on the vorticity transport theory they are related in a known manner. Measurements made at the National Physical Laboratory [58] show a large difference between the distributions of temperature and velocity and confirm the accuracy of the theoretical distributions given by the vorticity transport theory for the case of two-dimensional motion when the turbulence is confined to the plane of the mean motion. On the other hand, experiments made by Elias [62] show that the distributions of velocity and temperature in the turbulent boundary layer of a heated flat plate are very nearly identical, as would be expected from the momentum transfer theory: and calculations (unpublished) by Goldstein for the flow between long parallel plates show that the velocity distribution calculated from Kármán's formula for the apparent shearing stress obtained by the momentum transport theory is in closer agreement with the measured distribution than that calculated by the vorticity transport theory.

The mixing length l in the momentum and vorticity transport theories is considered only in relation to assumptions, more or less arbitrary, concerning the effect of turbulence on the mean motion, or of the mean motion on turbulence. It appears as a fictitious length, the value of which is found by observations of the distribution of mean velocity, temperature, etc. In Taylor's Statistical Theory of Turbulence [48] the scale of turbulence and its statistical properties in general are

given exact interpretation by considering the correlation between the velocities at various points of the field at one instant of time, or between the velocity of a particle at one instant of time and that of the same particle at some definite time, ξ, later. The diffusion of particles starting from a point ($y=0$) is shown to depend on the correlation R_ξ between the velocity of a particle at any instant and that of the same particle after an interval of time ξ. In continuous turbulent movements R_ξ is a function of ξ such that $R_\xi=1$ when $\xi=0$, and $R_\xi \to 0$ when ξ is large. If $\overline{Y^2}$ is the mean square of the distance through which the particles have diffused in time t it is proved that

$$\frac{1}{2}\frac{d}{dt}(\overline{Y^2})=\overline{Yv}=\overline{v^2}\int_0^t R_\xi d\xi.$$

When the irregular motion is such that it is possible to define a time T such that $R_\xi=0$ for all values of ξ greater than T, so that there is no correlation between the velocities of a particle at the beginning and end of the time interval T, then

$$\overline{Yv}=\overline{v^2}\int_0^T R_\xi . d\xi,$$

and it is possible to define a length l_1 such that

$$l_1=(\overline{v^2})^{\frac{1}{2}}\int_0^T R_\xi d\xi.$$

Taylor points out that this length l_1 bears no relation to any process of mixture, and that in this theory mixture has no effect on the diffusive power of turbulent motion.

The statistical properties of a field of turbulent flow when defined in the Eulerian manner are also investigated, and a length which represents in a definite way the scale of the turbulence is obtained. Taylor says that whatever ideas may be associated with the diameter of an eddy, a high degree of correlation must exist between the velocities at two points which are close together compared with this diameter, and the correlation is likely to be small between the velocity at two points situated many eddy diameters apart. The correlation R_y between the values of u at two points distant y apart in the direction of y is considered, and if R_y falls to zero at $y=Y$, then a length l_2 can be defined such that

$$l_2=\int_0^\infty R_y dy=\int_0^Y R_y dy.$$

This length l_2 is taken as a possible definition of the "average size of eddies", and it can be regarded as the analogue in the Eulerian system of l_1, which is defined in the Lagrangian system. The correlation coefficient $R_y=\overline{u_0 u_y}/\overline{u^2}$ can be determined from measurements of $\overline{u^2}$ and $\overline{u_0 u_y}$ with a compensated hot wire, and by introducing heat at a concentrated source or line source in an air stream and measuring the spreading of the heat behind the source, the quantity $\frac{d}{dt}(\overline{Y^2})$ can be calculated and

the value of $\int_0^t R_\xi d\xi$ for various values of t determined. Taylor's statistical theory has led to very valuable results. It is predicted and experimentally verified that when turbulence of a definite scale is produced in an air stream by means of a honeycomb or regular mesh (M), the turbulence dies down so that

$$\frac{U}{(\overline{u^2})^{\frac{1}{2}}} = \frac{Ax}{M} + B,$$

where A is a universal constant and B depends on the choice of the origin taken in the downstream co-ordinate x. Both l_1 and l_2 have been measured at the National Physical Laboratory, and it has been shown that they are directly proportional to M. It is also proved and experimentally verified that the rate of dissipation of energy, $\overline{W}$, is given by

$$\overline{W} = 15\mu\,(\overline{u^2}/\lambda^2),$$

where λ is the "average size of the smallest eddies" which are responsible for the dissipation of energy by viscosity. The theory has also been applied to determine the distribution of dissipation across a section of a parallel wall channel, and it is shown that in the region near the walls turbulent energy is produced more rapidly than it is dissipated, and that in the central region the reverse is the case. These results are subject to the restriction that the Reynolds' number of turbulence, $l(\overline{u^2})^{\frac{1}{2}}/\nu$, is greater than some number which must be determined by experiment.

§ 10. STATISTICAL MEASUREMENTS OF TURBULENCE

Records of the longitudinal speed fluctuations in a turbulent stream can be obtained with an Einthoven oscillograph used in conjunction with a hot wire. Such records are difficult to analyse statistically and recently a method has been developed by Simmons and Salter[63] at the National Physical Laboratory for determining the frequency distribution function $F(u)$ and the root-mean-square velocity $(\overline{u^2})^{\frac{1}{2}}$, by amplifying the out-of-balance potential variations generated across the Wheatstone Bridge containing a hot wire, and recording them by allowing the light from the mirror of a balanced-armature oscillograph to fluctuate to and fro across a stationary photographic plate for a period of 4 or 5 min. When developed, the plate reveals a dark image of variable density on a light background, and by measuring the distribution of this intensity photometrically, the total equivalent time of exposure at all points of the record is found, and from these measurements the velocity frequency curves are predicted. The method was applied to the analysis of the free stream turbulence in a small wind tunnel, and to the turbulent wake behind a grid composed of horizontal and vertical strips. It was found, for the condition of fully developed turbulence, that the frequency distribution closely resembled that of the normal error law, given by

$$F(u) = \frac{h}{\sqrt{\pi}}\, e^{-h^2u^2}.$$

Statistical measurements of the three components u, v, and w of the disturbed velocity for flow in a square pipe have been made by Townend[64], by a method which allows small masses of air to be identified and their motions observed. A succession of sparks is passed between very fine electrodes situated in the air stream, and the small hot masses of air are made visible by the Schlieren method, and cinematograph records of their motions taken as they travel downstream. The mean speed and the components of disturbed velocity are determined from these records. It was found that the maximum values of u, v and w were in good agreement with the values obtained from earlier experiments with an ultra-microscope[20] and also that the frequency curves for u, v, and w closely resembled those given by the normal error law. The maximum values of u, v and w were roughly three times the root-mean-square values, and about 3·75 times the average values, neglecting sign.

REFERENCES

(1) Dryden. *J. Wash. Acad. Sci.* **25**, No. 3 (1935).
(2) Simons and Brown. *Aero. Res. Comm., R. and M.* No. 1547 (1934).
(3) Kutta. *S.B. Akad. Wiss. Wien*, **41**, 65 (1911).
(4) Joukowski. *Bull. Inst. aérodyn. Kouchino*, Fasc. 1 (1906).
(5) *Aerodynamic Theory*. W. F. Durand, Editor-in-Chief: vol. **2**, Division E, "General Aerodynamic Theory—Perfect Fluids"; and vol. **4**, Division J, "Applied Aerofoil Theory".
(6) Bryant and Williams. *Philos. Trans.* A, **225**, 199 (1925).
(7) Tanner. *Aero. Res. Comm., R. and M.* No. 1353 (1930) and T. 3658 (1935).
(8) Lanchester. *Aerodynamics* (Constable and Co., Ltd., 1907).
(9) Lanchester. *Proc. Instn Auto. Engrs*, **9** (1915).
(10) Lanchester. *J.R. aero. Soc.* **30**, 587 (1926).
(11) Prandtl. *Z. Flugtech.* (February, 1912).
(12) Prandtl. *Nachr. Ges. Wiss. Göttingen*, I. Mit. (1918); II. Mit. (1919).
(13) Prandtl. *Rep. Amer. Nat. adv. Comm. Aero.* No. 116 (1921).
(14) Glauert. *The Elements of Aerofoil and Airscrew Theory* (Cambridge University Press, 1926).
(15) Fage and Simmons. *Philos. Trans.* A, **225**, 303 (1925).
(16) Caldwell and Fales. *Rep. Amer. Adv. Comm. Aero.* No. 83, **6**, 77 (1920).
(17) Piercy. *J.R. aero. Soc.* **27**, 488 (1923).
(18) Stanton, Marshall and Bryant. *Proc. roy. Soc.* A, **97**, 413 (1920).
(19) Fage and Falkner. *Proc. roy. Soc.* A, **129**, 378 (1930). Also *Aero. Res. Comm. R. and M.* No. 1369 (1931).
(20) Fage and Townend (with Appendix by Taylor). *Proc. roy. Soc.* A, **135**, 656 (1932).
(21) Fage. *Phil. Mag.* **21**, 80 (1936).
(22) Nikuradse. *Z. angew. Math. Mech.* **11**, 409 (1931).
(23) Prandtl and Schlichting. *Werft, Reed. Hafen*, **15**, 1 (1934).
(24) Prandtl. Math. Kong. Heidelberg (1904).
(25) Blasius. *Z. Math. Phys.* p. 1 (1908).
(26) Hiemenz. Inaugural Dissertation, Göttingen (1911).
(27) Kármán. *Z. angew. Math. Mech.* **1**, 235 (1921).
(28) Pohlhausen. *Z. angew. Math. Mech.* **1**, 261 (1921).
(29) Dryden. *Rep. Amer. Nat. Adv. Comm. Aero.* **20**, 435, No. 497 (1934).
(30) Howarth. *Aero. Res. Comm. R. and M.* No. 1632 (1934).
(31) Piercy and Preston. *Phil. Mag.* series 7, **21**, 995 (1936).
(32) Boltze. Inaugural Dissertation, Göttingen (1908).

(33) TOMOTIKA. *Aero. Res. Comm. R. and M.* No. 1678 (July 1935).
(34) HOWARTH. *Proc. roy. Soc.* A, **149**, 558 (1935).
(35) HELMHOLTZ. *Phil. Mag.* **36**, 337 (1868).
(36) KELVIN. *Phil. Mag.* **24**, 188 and 272 (1887).
(37) RAYLEIGH. *Collected Papers*, 1–6 (1879–1914).
(38) OSBORNE REYNOLDS. *Philos. Trans.* A, **186**, 123 (1895).
(39) ORR. *Proc. Roy. Irish Acad.* **27**, 9 (1907).
(40) TAYLOR. *Philos. Trans.* A, **223**, 289 (1923).
(41) SOUTHWELL and CHITTY. *Philos. Trans.* A, **229**, 205 (1930).
(42) NAUMANN. *Forschung*, No. 3 (1931), and No. 3 (1935).
(43) NIKURADSE. *Z. angew. Math. Mech.* No. 3 (June 1933).
(44) SCHILLER. *Proc. Int. Congr. Appl. Mech.*, Stockholm (1933).
(45) DRYDEN. *Rep. Amer. Nat. Adv. Comm. Aero.* **15**, 357, No. 320 (1929).
(46) MOCK and DRYDEN. *Rep. Amer. Nat. Adv. Comm. Aero.* **19**, 131 Rept. No. 448 (1933).
(47) ZEIGLER. *Proc. K. Akad. Wet. Amst.* **34** (1931).
(48) TAYLOR. *Proc. roy. Soc.* A, **151**, 421 (1935).
(49) SCHAUBAUER. *Rep. Amer. Nat. Adv. Comm. Aero.* No. 524 (1935).
(50) SIMMONS and BROWN. *Rep. Aero. Res. Comm.* T. 3219 (1932).
(51) RELF. *J.R. aero. Soc.* **39**, 1 (1935).
(52) OSBORNE REYNOLDS. *Philos. Trans.* A, **184**, 123 (1894).
(53) BURGERS. *Proc. Roy. Acad. Sci. Amst.* **32**, 414, 643, 818 (1929); **36** (1933).
(54) TRUBRIDGE. *Rep. Progr. Phys.* **1**, 43 (1934).
(55) PRANDTL. *Z. angew. Math. Mech.* **5**, 136 (1925).
(56) PRANDTL. *Verh. der* 3 *Int. Kong. Tech. Mech.*, Zürich, p. 62 (1926).
(57) PRANDTL. *Z. Ver. dtsch. Ing.* **77**, 105 (1933). [References to 35 papers given.]
(58) TAYLOR. *Proc. roy. Soc.* A, **135**, 685 (1932). [With Appendix, FAGE and FALKNER.]
(59) TAYLOR. *Philos. Trans.* A, **215**, 1 (1915).
(60) KÁRMÁN. *Gött. Nachr.* p. 58 (1930).
(61) KÁRMÁN. *Proc. Int. Congr. Appl. Mech.*, Stockholm, **1**, 85 (1930).
(62) ELIAS. *Abh. aerodyn. Inst. Aachen*, **9**, 10 (1930).
(63) SIMMONS and SALTER. *Proc. roy. Soc.* A, **145**, 212 (1934).
(64) TOWNEND. *Proc. roy. Soc.* A, **145**, 180 (1934).

THE UPPER ATMOSPHERE

§§ 1 to 11 by S. CHAPMAN

§§ 12 to 16 by S. CHAPMAN and W. C. PRICE

§ 1. INTRODUCTION

The principal data concerning the atmosphere at any level may be divided into two classes, physical and chemical. The chief among the former are the pressure, density, and temperature, and among the latter, the chemical constituents (including dissociated, ionized, and excited as well as normal forms) and their proportions. These data are neither all independent nor constant. We seek to know their values and interrelations, their changes and the processes which bring about their changes. Although for the lower part of the atmosphere, directly accessible to our instruments, we have fuller data than for higher regions, we are still far from understanding the changes that occur there, even at levels which seriously affect human life and society. The processes that go on in the upper levels are certainly not less complex, and, as we are less able to determine the principal data there, our understanding of the conditions is still very imperfect. Nevertheless it is possible to gain information about upper-atmospheric phenomena from a surprising number of different sources.

By skilful design it is possible to make a variety of light-weight physical instruments capable of being carried up to heights of at least 30 km. by unmanned balloons[1]; for purely scientific research this method seems more promising, as well as far more economical, than the use of manned balloons, with their smaller height range. Regener[2], for example, who has been specially successful in this work, has measured the temperature and pressure, obtained samples for test of composition, measured cosmic ray intensities, and taken photographs of the solar spectrum (throwing light on the ozone content) up to such heights. Moltchanoff, Väisälä and others have added very light radio transmitters to the instruments sent up by balloons[3]; by these the readings of the various instruments are automatically signalled and can be recorded on the ground at once, without waiting for the instruments to descend and be located and returned to the observatory that sends them up.

Much information concerning the upper atmosphere may also, however, be gained from observations made at ground level, as, for example, those on aurorae[4], meteors and their trails[5], and on high clouds, which are sometimes visible between 20 and 30 km.[6], and also (after dark) at 80–90 km.[7]. These are all phenomena that appear in definite regions of the atmosphere, which are visually recognizable, and can be located by trigonometric measurements (visual or photographic) from two or more points on the ground; the motion as well as the situation of the

aurorae, meteor trails, and high clouds can thus be found. Much further material for study is obtained by observing the spectrum of the light received.

The intensity, polarization and spectral composition of sunlight (direct or scattered), received from different directions at different solar altitudes, affords an almost infinitely rich material for study, containing much information concerning the atmosphere traversed, though in many ways difficult to interpret. The observation of moonlight and starlight permits this method of atmospheric investigation to be extended, in some degree, to the night hours; at which time also the self luminescence of the atmosphere (the light of the night sky, or non-polar aurora) has a fascinating story (still largely mysterious) to tell, concerning physical and chemical transformations occurring at high levels not yet definitely located.

Our resources do not stop short, however, at observation of phenomena that occur naturally; it is possible to project waves of various kinds into the high atmosphere and, by observing their return, to gain information which adds greatly to our knowledge. Electromagnetic or radio waves have proved of special value in this way; sound waves have also given useful results[8], and there is hope that ordinary light waves (searchlight beams) may likewise contribute new knowledge[9].

Further knowledge about the upper atmosphere may also be gained in other and somewhat unexpected ways; the transient magnetic variations observed at the earth's surface are mainly produced high in the earth's atmosphere[10], and furnish certain information, scarcely obtainable otherwise, as to the electrical conductivity and ionization there. The large-scale atmospheric oscillations, particularly those of half-daily period (solar[11] and lunar[12]) are also involved in these magnetic variations, and for this reason, and also independently, they contribute to our knowledge of atmospheric conditions at high levels.

Having thus briefly summarized the sources of our knowledge of the upper atmosphere, various topics will be discussed in more detail; but in view of the article by Prof. E. V. Appleton in vol. 2 of these *Reports*, on radio-exploration of upper-atmospheric ionization, little further on this subject will be mentioned here.

§ 2. THE INTER-RELATION BETWEEN PRESSURE, COMPOSITION AND TEMPERATURE

Though vertical movements in the atmosphere are of considerable importance, vertical accelerations are very small except in limited regions and for brief periods: hence the height-distribution of the atmospheric pressure p is effectively governed by the statical equation

$$\frac{dp}{dh} = -\rho g \qquad \text{.....(1),}$$

where h denotes the height, ρ the density and g the acceleration of gravity.

If the various atmospheric constituents at any height h have molecular masses $m_1, m_2, \ldots$, and if $n_1, n_2, \ldots$ molecules of each are present per unit volume, then

$$\rho = n_1 m_1 + n_2 m_2 + \ldots \qquad \text{.....(2).}$$

Also, if k denotes Boltzmann's constant ($=1{\cdot}372.10^{-16}$ in c.g.s. units) and T the absolute temperature,

$$p=k\,(n_1+n_2+\ldots)\,T \qquad \ldots\ldots(3).$$

It is convenient to write

$$n_0=n_1+n_2+\ldots \qquad \ldots\ldots(4),$$

$$n_0m_0=n_1m_1+n_2m_2+\ldots \qquad \ldots\ldots(5),$$

so that n_0 is the total number of molecules (including independent atoms and free electrons and ions) per c.c., and m_0 is the mean molecular mass. Then (2), (3) and (1) may be written in the equivalent forms

$$\rho=n_0m_0 \qquad \ldots\ldots(6),$$

$$p=kn_0T \qquad \ldots\ldots(7),$$

and

$$\frac{1}{p}\frac{dp}{dh}=-\frac{\rho g}{p}=-\frac{n_0m_0g}{kn_0T}=-\frac{m_0g}{kT}\equiv-\frac{1}{H},$$

or

$$\frac{d\log p}{dh}=-\frac{1}{H} \qquad \ldots\ldots(8),$$

where H is a quantity, having the dimensions of a length, defined by

$$H=kT/m_0g \qquad \ldots\ldots(9).$$

In an atmosphere which is throughout uniform in composition and temperature (so that m_0 and T are constant), and which is sufficiently restricted in height to permit g to be treated as constant (though really it decreases inversely as the square of the distance from the centre of the earth), the quantity H is itself constant. The equation (8) then has the simple solution

$$p=p_0e^{-h/H} \qquad \ldots\ldots(10),$$

that is, the pressure decreases upwards exponentially; p_0 here denotes the pressure at ground-level, where $h=0$. In these circumstances, also, $\rho\propto p$, and

$$\rho=\rho_0e^{-h/H} \qquad \ldots\ldots(11),$$

where ρ_0 is the density at ground-level. The total mass (M) of the atmosphere, per vertical column of unit cross-section, is then $\int_0^\infty \rho dh$ or ρ_0H, the same as for an atmosphere of *uniform* density and total height H. For this reason, and in reference to these special conditions, H is usually called the "height of the homogeneous atmosphere"

In the terrestrial and other atmospheres the quantity H, as defined by (9), varies with the height, and the simple solution (10) of the equation (8) is not valid. Nevertheless H remains of importance as a measure of the rate of *proportional* variation of the pressure with height; this may be illustrated by rewriting (8) in the form

$$\frac{d\log p}{d\,(h/H)}=-1,$$

appropriate to any level h; here H is to be regarded as a constant having the actual value, given by (9), appropriate to that level; H may then be interpreted as a unit

of height-measurement relative to which, at the given level, the rate of upward decrease of $\log p$ is unity. It is convenient to adopt a name for H in general; the term "height of the homogeneous atmosphere" is clearly not appropriate when H varies with height, and the name (local) "*scale-height*" may be suggested.

To know the height-distribution of the *pressure*, it is sufficient to know H as a function of h, and to solve (usually numerically or graphically) the equation

$$\frac{dp}{dh} = -\frac{1}{H} \qquad \text{.....(12)}$$

or

$$p_0 - p = \int_0^h \frac{dh}{H} \qquad \text{.....(13)};$$

it is *not* necessary for this purpose to know the temperature and composition separately. Conversely a knowledge of the height-distribution of pressure would enable H to be calculated as a function of the height; this would indicate the height-distribution of m_0/T, since k and g are known.

If both p and T are known as functions of the height, H can be calculated from p, yielding the value of m_0/T, whence, knowing T, we can find the value of m_0; also from (7), knowing p and T, we can derive n_0, so that from (6) we can calculate the density ρ at all heights.

§ 3. APPLICATION TO BAROTHERMOGRAPH READINGS

Regular meteorological practice in the measurement of upper-air data includes the sending up, by means of small balloons, of an instrument called a barothermograph; this provides a curve of which the two coordinates at any point represent simultaneous values of p and T at some unknown height h. These records are interpreted on the assumption that m_0 is the same, at all heights traversed, as at the ground-level; the equation (8) can in fact be re-written—using also (9)—

$$dh = -\frac{k}{m_0 g}\frac{T}{p}\,dp,$$

so that, if m_0 is a constant,

$$h = \frac{k}{m_0 g}\int_p^{p_0} \frac{T}{p}\,dp \qquad \text{.....(14)};$$

the quantity on the right can be deduced from the T, p curve, and so gives the height h to which each point of the curve corresponds.

The assumption that m_0 is constant has recently been verified, within amply adequate limits compared with the other sources of error in (14), by the analysis of samples of air from levels up to almost the greatest heights yet attained by the small balloons used. This justifies the existing method of interpreting the barothermograph records. But if, in the future, technical advances enable such T, p records to be extended up to heights of 100 km. or more, it is fairly safe to say that variations in m_0 will have to be taken into account in (14); that is to say, m_0 will have to be placed within, and not outside, the integral sign.

§ 4. THE COMPOSITION OF THE TROPOSPHERE

The water-vapour content of the troposphere (the region up to about 10 km. height in which T decreases upwards) is very variable, owing to processes of evaporation, condensation, precipitation, convection, and diffusion; its amount may approach 4 per cent by volume in the lower, warmer levels.

In the region (the stratosphere) above the troposphere, where T remains nearly uniform with respect to height up to at least 30 km., T is so low (about 219° abs.) that even the saturation pressure of water vapour is extremely small (corresponding to a density of about $2{\cdot}5 \cdot 10^{-8}$ g./cm.^{3}

In discussing the chemical composition of the atmosphere, it is usual to consider only the proportions of the different constituents of the *dry* air, separately discussing the humidity, or the proportion of water vapour to the total of the dry constituents.

Analysis of samples of air from various geographical regions and at different heights in the troposphere show a remarkable degree of constancy in the composition of the dry air. Table 1 gives the composition according to the best data now available[13].

Table 1. Composition of the troposphere (per cent by volume)

Nitrogen	78·08	99·99
Oxygen	20·95	
Argon	0·93	
Carbon dioxide	0·03	
Neon	$1{\cdot}8 \times 10^{-3}$	
Helium	5×10^{-4}	
Krypton	1×10^{-4}	
Xenon	1×10^{-5}	
Ozone	Variable: $> 1 \times 10^{-6}$	
Radon (average near ground)	6×10^{-18}	
Hydrogen	Doubtful: $< 1 \times 10^{-3}$	

§ 5. MIXING IN THE TROPOSPHERE AND STRATOSPHERE

The value of m_0, if water vapour is neglected, depends mainly on the nitrogen and oxygen; it corresponds to a molecular weight (in the chemical sense) of 28·97. Since the molecular weight of water vapour is 18, the mean molecular weight of the actual moist air of the troposphere may be taken as being between 28·97 (dry air) and 28·55 (air containing 4 per cent water vapour).

Other variable constituents of the air in the troposphere are *radon*—which is emitted from the ground, and decays with a half-life of 3–8 days, so that it is not found above a height of 4 km., nor over the oceans; *ozone*, which appears to be formed in the stratosphere and to be carried down by convection and diffusion into the troposphere, there to disappear slowly by chemical processes not yet definitely ascertained; and *carbon dioxide*, which in badly ventilated places may vary considerably from the mean percentage (0·03), but which in the open air deviates very little from the mean. These variations have scarcely any effect on m_0.

The most sensitive chemical and spectroscopic tests, by Claude in Paris, Krogh in Copenhagen, and Günther in Königsberg, have failed to detect any hydrogen

as a constant constituent of tropospheric air; its amount is certainly less than 0·001 per cent.

The composition of the troposphere, apart from the non-permanent constituents, is maintained uniform within the present limits of experimental determination by the constant mixing effected by winds and turbulence. This mixing of different layers is also responsible for the upward decrease of temperature in the troposphere at the rate of about 5° C. per km.; this is distinctly less than the adiabatic rate of upward decrease of T (10° C. per km.), showing that other influences (conduction and radiation) tend to promote a more uniform temperature distribution. In the accessible part of the stratosphere such influences on the temperature clearly predominate over convection. This does not imply that convective mixing of the air is there totally absent, but only that the influence of mixing on the *temperature* is there far less, relatively to the influences promoting uniformity, than in the troposphere. Observations of the motion of the mother-of-pearl clouds sometimes seen at 20–25 km. height in the stratosphere, and of luminous night clouds at about 80 km. height, as also of meteor trails, show that there are at times considerable wind velocities in the upper atmosphere, and that these vary both with time and height. Such winds must promote some degree of mixing of the air. The variations in the ozone content of the stratosphere, which are associated with weather changes in the troposphere, also suggest the same conclusion. It seems likely, however, that the mixing of the air in the stratosphere will be less powerful than in the troposphere, and that some scope will there be given to the natural tendency of the constituents of different molecular weight to distribute themselves according to Dalton's law of diffusive equilibrium. This should show itself in an upward decrease of the volume percentage of the heavier relative to the lighter constituents. In particular, oxygen should *decrease* upwards relative to nitrogen, and the helium content should *increase* upwards. Ultimately, at great heights, the lightest constituents should predominate over all others.

§ 6. CALCULATIONS OF UPPER ATMOSPHERIC DISTRIBUTION

Various authors[14] have in the past made calculations of the change of composition of the atmosphere with height, according to Dalton's law, as if the uniform height-distribution of T in the stratosphere, observed from about 10 km. height to about 30 km., implied a sudden transition at the base of the stratosphere from a state of complete mixing to one of complete diffusive equilibrium. Such calculations involve only a knowledge of the composition of the air at the base of the stratosphere, and a knowledge of the temperature at all heights considered. As regards the latter, the simplest assumption was to take T as the same at all heights above 10 km.; any other assumption would not alter the general tendency of the lighter constituents to increase upwards relative to the heavier ones, though it would alter the actual calculated distribution; an upward rise in T would reduce the rate at which the air density as a whole decreased upwards.

6·1. *The hydrogen content.* Such calculations naturally led to the conclusion

that the air above a certain height would be composed almost entirely of the lightest constituent, formerly assumed to be hydrogen; the calculated level at which hydrogen became predominant depended on the assumed percentage amount at the base of the stratosphere. Since, as has been stated, this is too small to be determined accurately, no reliance can be placed on the hydrogen densities calculated. There is, indeed, much reason to think that at no height in the atmosphere is hydrogen an important constituent. The upper atmosphere is excited to luminescence both regularly—giving the light of the night sky, which appears to be due to recombination processes, following upon the dissociative action of sunlight on the air during the hours of daylight—and spasmodically during auroral displays; the spectrum of the light emitted in either case appears on analysis to be due mainly to oxygen and nitrogen; no lines or bands due to *hydrogen* have yet been identified with certainty. This negative evidence suggests the absence of hydrogen from the high atmosphere.

6·2. *Escape of gases into space.* While this evidence by itself may not be conclusive in showing that hydrogen is absent, supporting evidence is afforded if we accept the view, now increasingly held, that the outer ionized layers of the atmosphere are at temperatures of the order of 1000° abs. for an appreciable fraction of the time. At such temperatures the rate of loss of hydrogen molecules, owing to their high thermal velocities, by escape into free space, would be considerable[15], and in a few million years (a short time compared with the age of the earth) all the hydrogen originally present in our atmosphere would have been lost. The escape would be much faster still if, as seems reasonable, an appreciable proportion of the hydrogen on the confines of the atmosphere were dissociated into atoms (having thermal velocities 1·4 times as great as hydrogen molecules) by absorption of sunlight of very short wave-length. Hence the possibility that hydrogen is practically absent not only from the troposphere but throughout the whole atmosphere deserves serious consideration.

6·3. *The helium content.* If hydrogen be absent from the high atmosphere, the uppermost layers should consist almost entirely of the next lightest constituent, helium. But spectroscopic indications of the presence of helium are as lacking as those for hydrogen. If the temperature of the high atmosphere attains 1000° abs. for an appreciable fraction of each year, helium also should escape at a rate which would imply the disappearance of any original helium layer, in a short time compared with the age of the earth. Helium, however, is continually being added to the atmosphere at its base, owing to radio-active processes within the earth's crust. The existence of a steady state implies the escape of helium from the top of the atmosphere at the same rate as replenishment occurs at the base. The latter rate is of course not well determined, but an equal rate of escape at the top of the atmosphere would seem quite feasible if temperatures of the order of 1000° occur at high levels[16]. The existence of a steady state implies also a steady upward flow at the same rate at every level in the atmosphere. The amount of helium present at each level, and slowly ascending at the required rate, has not yet been calculated theoretically. It has, however, been established, by recent work of Paneth and

Glückauf[17], with the aid of L. H. G. Dines[18], that up to 18 km. the helium ratio in the stratospheric air is the same as near the ground, but that at certain times a definite small increase of helium content is found at 21 km., the largest increase being by 8 per cent of the normal ratio. This appears to indicate that, at least under certain conditions (when presumably a layer of air has remained nearly static for a longer time than usual), an approach towards diffusive equilibrium of the helium can occur; calculation shows that in a layer of completely static air the ordinary helium ratio should increase upwards by 1 per cent in the first 75 m. The results of Paneth and Glückauf suggest that the diffusive separation of helium is distinctly less effective than this.

6·4. *The oxygen content.* Regener[19] has recently announced the results of analyses of the oxygen content of stratospheric air drawn from heights up to 28 km. Up to about 20 km., just as for helium, no appreciable change was observable, but samples taken from above 20 km. showed a small diminution of the oxygen ratio, the deficit at the highest level (28 km.) being 2·5 per cent of the normal ratio. The values appear to depend not on the height only, and Regener suggests that they are connected with weather processes; in particular, that air coming from the polar stratosphere probably shows more evidence of diffusive separation, owing to the existence of more stable conditions there, than air from lower latitudes.

6·5. *The convection-diffusion interface at height h_{cd}.* These important researches supply a valuable basis of fact for the conclusion that some degree of diffusive separation of the atmospheric constituents obtains at heights above 20 km. No sharp transition between regions of no diffusion and regions of complete diffusive separation is to be expected, but for purposes of numerical illustration it is useful to calculate the distribution of the various constituents, on the hypothesis that such a sharp transition *does* occur at some level[20]. This level may be called the Convection-Diffusion interface, at height h_{cd}. To give suitable results for the higher regions, the height h_{cd} should be taken somewhat above the level, 20 km., where are found the first signs of diffusive separation, there only partial.

If h_{cd} be taken as 30 km., calculation shows that the helium mass density* would not exceed the nitrogen mass density till somewhat above 140 km., if $T = 220°$ over this height range.

If in intermediate levels T exceeds or falls short of this value, the He/N_2 mass-density ratio would become unity at a correspondingly higher or lower level; at the same level the O_2 mass density would be only about one-thirtieth of the N_2 density.

It must not be forgotten, however, that though partial diffusive separation may proceed for some distance from 20 km. upwards, the process may be arrested in some higher layer should the conditions there favour an upward increase of turbulence and mixing.

6·6. *Layers of greater or less convective stability.* Turbulence depends on the presence of causes tending to produce relative motions of adjacent masses of air; but the effect of these causes in producing vertical interchanges will depend on the value of the vertical temperature gradient. A quickly ascending or descending

* Calculated statically, not allowing for any upward flow of helium.

mass of air is cooled by expansion or heated by compression; if the change of volume occurs adiabatically, the air will have the same temperature as its new surroundings, provided that the temperature gradient is 10° C. per km. This adiabatic gradient may be regarded as corresponding to neutral stability for rapid vertical interchanges. In the troposphere there is an upward decrease of temperature, but at only about half the adiabatic rate, so that ascending air becomes cooler and therefore heavier than the surrounding air at its new level, and conversely for descending air. In these circumstances vertical interchanges of air are discouraged, though not prevented. In the constant-temperature region of the stratosphere this discouragement is much greater, though even in that region differential motions of air do occur. There is reason for thinking that T may actually increase upwards from a height of about 30 to about 50 km.; it may be hoped that actual barothermograph records will in the not distant future remove any doubt as to the facts; if T does increase upwards in this region, any causes tending to promote vertical interchanges will there be held specially severely in check. But there is some reason to suppose that above this layer, again, the temperature decreases once more with increasing height[11, 49]; this would render the conditions there relatively less unfavourable for mixing, perhaps no more so than in the troposphere (though the causes actually promoting mixture are probably greater in the troposphere than anywhere higher up). Hence while the helium and oxygen ratios may vary with height, especially in the region 30–50 km., their further separation may be arrested in a layer extending upwards for some distance above 50 km. This would naturally affect such estimates as those given in § 6·5.

§ 7. SOUND PROPAGATION AND UPPER-AIR TEMPERATURE

Evidence for an increase of temperature beginning at some height not much above 30 km. is afforded by observations of the abnormal propagation of sound to great distances[8]. The velocity of sound in a gas is proportional to the mean thermal molecular velocity in the gas, which itself is proportional to $\sqrt{(T/m_0)}$. Sounds generated near the ground are audible with decreasing intensity at increasing distance from the source; hence an inner zone of audibility is surrounded by a region of inaudibility. If the sound is loud enough, it may again be heard "abnormally" in a further zone surrounding the first region of inaudibility. The inner radius of this outer zone of abnormal audibility is of the order 100 km. Observations show that the sound received there is travelling obliquely downward; hence it is ascribed to abnormal propagation in which the sound travels upwards to some height in the atmosphere, whence its path is bent downwards again. In this region where the sound-path is concave relative to the ground, the upper parts of the wave-fronts must be travelling faster than the lower parts; this implies an upward increase in the sound velocity, and therefore in $\sqrt{(T/m_0)}$ or T/m_0, in that region. The sound observations do not afford a very definite estimate of the height attained by the sound waves, and their interpretation is rendered more difficult

by uncertainty as to the wind velocity at those heights; but the height of the top of the sound-path may be taken as below 60 km. If m_0 is supposed constant, the increase in T/m_0 inferred from the sound observations would correspond to an increase of T to 300° abs. or more. In the past it has been suggested that, on the contrary, T remains constant and m_0 decreases, the decrease being attributed to the presence of a large proportion of hydrogen in the region concerned; this idea, never very probable, can now be definitely dismissed in view of the very slight change in m_0 up to 28 km. height, as indicated by Regener's samples. The fullest diffusive separation of hydrogen from 28 km. upwards could not produce the increase in its proportions at 60 km. necessary to explain in this way the increase of T/m_0 as due to a change in m_0.

In recent years the increase in T/m_0 has generally been attributed to a change in T. Very recently, however, Eropkin[21] has revived, in a different form, the suggestion that it is due to a change in m_0. Eropkin states that he has made new calculations of the ozone-oxygen equilibrium[50] which indicate practically complete dissociation of oxygen into the atomic form from 35 km. upwards. Considering only the nitrogen-oxygen content of the atmosphere, and assuming that their mass-ratio there retains the same value (3 to 1 approximately) as near the ground, the dissociation of all the oxygen molecules into atoms would reduce m_0 from 29 to about 24; this would have the same effect on T/m_0 as an increase of T from 220 to 266° (m_0 remaining constant). Such a change in T/m_0 seems insufficient to explain the sound observations, which therefore appear to imply *some* increase in T up to 60 km. or so, though the deflection of sound waves back to the ground may also be partly due to a decrease in m_0. The uncertainty emphasizes the need to extend balloon exploration of the atmosphere beyond the present limit of about 30 km., to determine both T and, if possible, m_0, above those levels.

§ 8. THE TEMPERATURE ABOVE 50 OR 60 KM.

Information as to the temperature or scale-height (H) of the atmosphere above 60 km. is still more difficult to obtain; but any method that will indicate the *density* of the atmosphere at a definite (high) level throws some light on the mean value of H between that level and 30 km. Lindemann and Dobson developed a theory of meteors[5] which, applied to the available data on the frequency of meteor appearance and disappearance at different heights, led them to conclude that the air density at, say, 100 km. much exceeds that which would be found there if T/m_0 were constant throughout the stratosphere ($T=220°$, $m_0=29$); they inferred that at some level in the stratosphere T must rise considerably above 220°, as the abnormal sound observations likewise indicate (though the latter give no information as to levels above about 60 km.).

8·1. *The world-wide atmospheric oscillations.* Another somewhat recondite source of information as to the value of H at high levels is afforded by the theory of the world-wide atmospheric oscillations, particularly the main oscillation of 12-hour period. On account of its regularity and large magnitude this is a meteoro-

logical phenomenon of outstanding interest theoretically (though of no immediate importance for such practical problems as weather forecasting); its presence strikes the eye at once on examining the records of any barograph from tropical latitudes. Its origin is now reasonably well accounted for as due in about equal measure to the sun's tidal and thermal influences, which together determine the phase. Its large magnitude, however, can only be explained as due to resonance—the near coincidence (within 3 or 4 min.) of the "forced" period of 12 hours with a free period of oscillation of the atmosphere, of the observed form. The atmosphere has more than one mode of free oscillation of the type which, so far as concerns the ground observations (of barometric pressure), is observed; the simplest such mode has a free period of about $10\frac{1}{2}$ hours, differing greatly from 12 hours. In the next simplest mode the associated atmospheric currents are not (as in the first mode) in the same phase at all heights, but are reversed above a certain height. The period of this mode depends on the height-distribution of H or T/m_0, which is known only up to 30 km.; the period can be calculated on the basis of different assumptions as to the values of H above this level. Pekeris[11] finds that the free period has a value just under 12 hours if, and only if, it is assumed that H first increases (as the sound observations suggest), and then decreases again. He gets the appropriate free period for an atmosphere in which T in the stratosphere increases from 220 to 370° abs. between 37 and 60 km. and then *decreases* again to 220° at 80 km., remaining constant for many km. above that level. Though recondite, this inference seems to have considerable weight and importance. It is a satisfactory feature of the theory that it implies a reversal of phase (above about 30 km.) of the atmospheric movements associated with the oscillation. This fits in with indications to the same effect derived from the daily magnetic variations, themselves due to the dynamo action of these atmospheric movements in the high ionized layers. The theory that the twelve-hourly oscillation is of the second mode also implies that the associated atmospheric velocity at high levels is not only opposite to that at the ground-level, but also much greater in magnitude; this removes a difficulty in the theory of the daily magnetic variations, which depend on the product of the atmospheric velocity and the electrical conductivity (or ionization) of the layers concerned. So long as the oscillatory movements are supposed to have the same velocity at all heights, the theory of the daily magnetic variation seemed to demand an intensity of ionization greater than would accord with the measures of this ionization afforded by radio methods; the greater velocity at high levels, according to the revised theory of the atmospheric oscillation, makes smaller demands on the ionization, and so affords a more favourable interpretation of the whole set of facts, barometric and magnetic.

8·2. *High-level temperatures from radio, auroral and other data.* Radio measurements provide another source of information as to the density of the atmosphere at certain great heights, because they indicate the value of the collision frequency of electrons in the ionized layers, and this frequency depends rather directly on the density. They also afford independent information as to the value of H, through the fact that H largely governs the thickness of the ionized layers, which can be

estimated from certain radio observations. Again, the daily movements of the upper ionized layers suggest large daily changes of temperature. From consideration of such data, and also of the radio echoes of long delay, Fuchs[22], for example, concludes that m_0 can at no level correspond to a predominantly hydrogen or helium atmosphere, and that in the higher ionized layers (F_1 and F_2) the temperature may rise as high as 1000 and 1900° respectively. Appleton's estimates of these temperatures are about half as great[23]. As regards the E layer, at about 100 km. it is so thin, according to Appleton[49], as to suggest a low temperature there of 200° abs. or even less. Vegard and Tønsberg[24], from a study of the intensity distribution in certain spectral bands of auroral light, likewise infer low temperatures (200–250° abs.) at about 100 km. height in the auroral region.

The time has not yet come when a self-consistent synthesis of the indications afforded by these various means can be arrived at; and if these indications do not all appear accordant at present, this is perhaps only to be expected, owing to uncertainties remaining in the interpretation of the data. It is highly desirable to develop and improve these less direct ways of obtaining information about the upper atmosphere; but it is at least equally desirable to strive to extend the height-range of simpler and more direct methods of upper-air investigation, particularly by balloons carrying barothermographs and sampling vessels. Another fairly direct method which promises well, but has not yet been applied in practice, is the measurement of the brightness at different points along a searchlight beam when the intensity of illumination is made to undergo a periodic change: this method, developed by Tuve, Johnson and Wulf from a suggestion by Synge[9], seems capable of giving information as to the scattering power, and hence the density, of the atmosphere up to at least 60 km.

§ 9. THE UNCERTAINTY AS TO CONVECTION AT GREAT HEIGHTS

The preceding discussion will have shown the difficulty attaching to calculations of the pressure, density and composition of the air at different heights, because of uncertainties as to the temperature distribution above 30 km., the degree of diffusive separation of the constituents at different levels, and the presence or absence of helium at great heights. As regards diffusive separation, the rapidity of the process is inversely proportional to the air density[25], and therefore it should play an increasingly important part as we ascend in the atmosphere. But unfortunately we are as yet quite unable to judge the intensity of the factors that, by promoting turbulence and mixing[26], counteract the tendency towards separation. It is certainly a striking fact that the auroral spectrum shows the presence of molecular nitrogen and atomic oxygen up to very great heights.

§ 10. CHEMICAL AND PHYSICAL TRANSFORMATIONS IN THE UPPER ATMOSPHERE

There are, however, still further and perhaps greater difficulties obstructing the calculation of densities and composition at great heights, arising from the occurrence of energetic chemical and physical transformations there, affecting

constituents that may in the lower atmosphere be treated as permanent and not subject to change. The chief transformations appear to be: the dissociation of molecular into atomic oxygen, which then, by recombination with other oxygen molecules, may form ozone; the ionization of molecular nitrogen and also, doubtless, of oxygen; and the production of various excited states of nitrogen and oxygen molecules or atoms. Such changes greatly complicate the already difficult problem of the constitution of the upper atmosphere.

These energetic processes in the upper atmosphere are brought to our notice in several ways: magnetic and radio data indicate the occurrence of the most energetic process of all, ionization, in the upper atmosphere. The interpretation of the auroral and night-sky spectra makes it clear that neutral but excited atomic oxygen is regularly present in the upper atmosphere, at least by night; and the theoretical study of the observations on atmospheric ozone led to the recognition that above a certain height in the atmosphere (not yet well determined) the oxygen is almost completely dissociated, existing almost entirely in the atomic form.

The dissociation, excitation and ionization of the atmospheric constituents, and the chemical and radiative processes which follow from them, occur chiefly in the upper atmosphere. There can be no doubt that this is because the first cause in every case is the entry into the earth's atmosphere, from without, of some radiation, of wave or corpuscular character, which becomes absorbed before reaching the ground. There are two radiations, however, which transform atmospheric constituents right down to ground-level; the most important is the cosmic radiation, which has extraordinary energy and penetration; though its impact on the air has results of intense physical interest, these results apparently do not much affect the main atmospheric processes. The other transforming and penetrating radiation is the ordinary visible radiation in the Chappuis band, which seems capable of dissociating the not very stable ozone molecules, but is only weakly absorbed by them.

§ 11. CORPUSCULAR AND WAVE RADIATIONS FROM THE SUN

The cosmic radiation seems to come from all parts of space, but the radiations which chiefly govern atmospheric processes come from the sun; they include sunlight (visible and invisible) and also corpuscular radiations, such as produce aurorae. The corpuscular character of the latter radiations is inferred from their peculiar geographical incidence: instead of falling only on the sunlit hemisphere, they enter the atmosphere over the dark half of the earth (as well as—almost certainly, though invisibly—over the sunlit half), and specially in polar regions, in and near a zone of about 20° diameter centred round each pole of the earth's magnetic axis; the only known explanation of this fact is that these radiations are electrically charged, and are guided polewards by the earth's magnetic field—which is known also to affect the primary cosmic radiation in a similar manner.

The *solar* origin of the corpuscular (auroral) radiations is attested by the incidence of aurorae in time, which shows daily, seasonal, annual and sunspot-cycle variations, and also a recurrence-tendency connected with the sun's rotation.

It is possible, and even likely, that the sun's corpuscular radiations also include uncharged particles that are undeflected by the earth's magnetic field, and hence fall, like sunlight, only on the sunlit hemisphere of the earth.

The speed of travel of the charged corpuscular radiations from the sun is not known. Estimates have been made on the basis of supposed connections of occurrences on the sun (such as the passage of notable sunspots across the sun's central meridian) with subsequent terrestrial phenomena, chiefly aurorae and magnetic storms; the mean interval found in such cases is from one to four days, corresponding to a speed of the order 1000 to 250 miles a second. These speeds seem too low to account for the penetration of the corpuscles into the earth's atmosphere, supposing that auroral rays are direct manifestations of the action of the solar corpuscles, *in situ*; because auroral rays habitually extend down to 100 km. height, and have been found on occasion to come down even below 70 km. Estimates of the speed of the particles from these penetration data must remain uncertain, however, until (*a*) the density distribution in the upper atmosphere is better known, and (*b*) until the mass of the solar corpuscles can be determined[30]. It is not yet known whether the corpuscles that produce the visible aurorae are electrons or particles of atomic mass.

The daily variation of the ionization in each of the main layers (E, F_1, F_2), which decay nightly, to be renewed after dawn each day, shows clearly that the cause of the ionization is some radiation from the sun that falls almost entirely on the sunlit hemisphere. It has been suggested that uncharged corpuscles from the sun play a part, along with sunlight (in the ultra-violet region of the spectrum), in this ionizing process.

Solar eclipses can throw light on this question, at least if the corpuscular speeds are notably less than that of light; for in that case the "hole" or shadow region carved by the moon in the beam of corpuscles will make an appreciable angle with the axis of the optical shadow, and two distinct eclipses (corpuscular and optical) will result. The former lies to the east of the ordinary (optical) eclipse, and occurs earlier, the interval between the two being of the order of two hours if the corpuscular speed is 1000 miles a second. Recent radio eclipse data, which indicate a marked decrease in the ionization of each of the principal ionized layers during "optical" eclipses, show that the main ionizing agent in the E and F_1 regions is certainly one that comes from the sun with the speed of light or very nearly this speed: hence it is natural to conclude that ultra-violet sunlight is responsible for the formation of these two ionized layers. The eclipse facts about the highest, F_2, layer, are less clear, but it seems not unlikely that there also sunlight is the main regular ionizing agent. The daily renewal of the ionized layers by sunlight is probably much the most energetic action of the sunlight on the upper atmosphere, and has many interesting consequences.

§ 12. THE ABSORPTION OF SOLAR RADIATIONS

Conditions in the high atmosphere seem to be governed largely by the difference between the absorptive powers of nitrogen and oxygen for sunlight of wave-length

shorter than about 3000° A.; nearly all this radiation is absorbed before it penetrates to ground-level. A knowledge of the nature of the absorption of this light by the different atmospheric constituents is very necessary if one is to understand the resulting phenomena. For this information we must rely on the work of spectroscopists, which is extensive but still very inadequate for the needs of upper-atmospheric physics. Their work indicates the strength of the absorption, that is, the number of cm. of a particular constituent, at N.T.P., needed for the practically complete absorption of light of a particular wave-length. The analysis of their spectra also affords information as to the effect of the absorption on the absorbing molecules. An atom may be excited or ionized or both; the excited state may be of short or long life—in the latter case the state is called metastable. Molecules on absorption may be transformed into (*a*) a stable excited molecular state, or (*b*) into an unstable excited state, leading rapidly to dissociation, into excited or unexcited atoms, or (*c*) into an ionized state, possibly also excited.

12·1. *Nitrogen absorption.* Molecular nitrogen (N_2) is relatively transparent in the extreme ultra-violet, and transmits freely from the visible region down to 1450 A. Here a progression of narrow discrete absorption bands occurs, which absorb small ranges of the spectrum in the region 1450–1000 A.[27] The absorption in the centre of these bands is not complete until the light has traversed a layer equivalent to 40 cm. of the gas at N.T.P. This absorption is much weaker than that of O_2 in the same spectral region. The sharp discrete nature of these N_2 bands indicates that the absorption excites the molecules to stable excited states, and does not cause it to dissociate into atoms.

No published data are available regarding N_2 absorption in the region 1000–800 A., but recent work (by W. C. Price) has shown that here also N_2 possesses sharp absorption bands; these are much stronger than those first mentioned, absorption being almost complete in a path-length of only a few tenths of a mm. of N_2 at N.T.P. A continuous region of absorption appearing at about 800 A. may possibly be correlated with the photo-ionization of the molecule, liberating an electron and leaving N_2^+ in its lowest state (the ionization potential is 15·5 e.v., corresponding to $\lambda = 795$ A.); but further work is needed before this can be stated with certainty.

In the region 800–650 A. Hopfield[28] has discovered a set of bands which form a molecular Rydberg series converging to a limit which corresponds to one of the excited states of N_2^+; in fact photo-ionization occurs on absorption of light of wave-length less than 661 A. (18·67 e.v.). The existence in the upper atmosphere of such an excited molecular nitrogen ion would involve the emission of light in the "negative" nitrogen bands, corresponding to the reversion of the molecular ion from its excited to its normal state. Slipher[29] and others have observed the presence of these negative bands of nitrogen in the spectrum of the light of the night sky within about an hour after sunset and an hour before dawn, but not during the rest of the night. This seems to indicate that the sunlight is at those times producing excited nitrogen ions in the high regions of the atmosphere, which are exposed to sunlight even when the sun is considerably below the ground horizon;

but that during the night this ionization and excitation ceases (except when corpuscles enter from outside, producing aurorae, in the spectrum of which the negative bands of nitrogen form a prominent feature).

The absorption coefficient of nitrogen for light in the region of the Hopfield bands is much greater than for the region 1000–800 A., the bands appearing when the light traverses only a few hundredths of a mm. of nitrogen at N.T.P.

Thus the available data on the far ultra-violet spectrum of N_2 indicates that absorption on the short-wave side of 795 A. (15·5 e.v.) and of 661 A. (18·67 e.v.) will cause photo-ionization. Excitation without ionization is produced by absorption of light in the region 800–1000 A. But nitrogen appears not to be dissociated by absorption of light in any of these bands. These bands are absorbed by molecular nitrogen in the ground state. The excited states of N_2 which lead to emission of light whose band spectrum is diffuse, indicating a state of predissociation, are not formed directly by absorption of light from the ground state. Nitrogen molecules in such excited states may be produced in the atmosphere by absorption of light by nitrogen molecules already excited by absorption (say for $\lambda > 800$) from the ground state, and these doubly excited nitrogen molecules may then dissociate, giving rise to nitrogen atoms; but this double process must occur relatively rarely, because most of the molecules in the first state of excitation will revert to the ground state very quickly, before they have time to absorb another light quantum. Hence atomic nitrogen may be expected to be present in the upper atmosphere, if at all, only in very small amounts. This accords with the fact that as yet we have no indication of the presence of atomic nitrogen in the atmosphere, either from the night-sky or from the auroral spectrum.

12·2. *Oxygen absorption.* The absorption of oxygen for light of short wave-length differs in almost every feature from that of nitrogen, and the difference has far-reaching consequences for the upper atmosphere.

Proceeding from the visible region towards short wave-lengths, the first ultra-violet absorption bands of O_2 that are encountered are the Herzberg bands[30], in the region 2400–2600 A. The light in this region is very weakly absorbed, and the bands appear in the spectrum only after the light has passed through long columns of oxygen at N.T.P. Light in this same region of the spectrum is however very strongly absorbed by an allotropic form of oxygen, namely ozone, which exists in the atmosphere, to the extent of 2–4 c.mm., at N.T.P., per sq.mm. column of air from the ground to the top of the atmosphere; this quantity is much more than sufficient to absorb this light completely, and there is little doubt that in the atmosphere this light is absorbed mainly by ozone and not by O_2.

The next shorter-wave set of absorption bands is the Schumann-Runge set[31], in the region 1925–1760 A. These appear strongly in the spectrum of light that has passed through a path of 2 or 3 cm. of O_2 at N.T.P. They form a progression of gradually converging bands, corresponding to increased vibration of the excited molecule formed by absorption of the light; and in the limiting state of the progression, corresponding to absorption of light of wave-length 1760 A., the molecule dissociates. The products of the dissociation are a normal O atom (^1P) and an

O atom in the metastable ^1D excited state. The free spontaneous passage of the latter to the ground state causes the emission of the "forbidden" red lines observed in the auroral spectrum.

Flory[32] contends that oxygen is dissociated not only at the limit (1760 A.) of the Schumann bands, but also, by pre-dissociation, by absorption in the region from 1950 A. to 1760 A.; the atoms formed in this case are both in the normal state ^3P, so that no atomic emission of light follows.

Proceeding further along the wave-length scale, we traverse a region in which O_2 exerts a very much stronger diffuse continuous absorption, extending from about 1760 to 1350 A.[33] Light in this region of the spectrum is completely absorbed by only 0·1 mm. of O_2 at N.T.P. The continuous character of the absorption spectrum indicates that the absorbing molecules become dissociated, and the product must be two O atoms, one in the normal ^3P state, the other in the excited ^1D state (these atoms possessing also, in the form of kinetic energy, a balance of the energy drawn from the light quantum). Complete absorption of sunlight in this region occurs after the light has passed through a small amount of O_2 which absorbs practically none of the light in the Herzberg region. The former light is therefore absorbed very high up, and will have the effect of maintaining the oxygen in the outer layers of the atmosphere almost entirely in the atomic form. The light in the Herzberg region penetrates far down into the stratosphere, though it does not reach the troposphere. Only a small proportion of the oxygen molecules in the lower part of the stratosphere will be dissociated; the O atoms produced there will mainly combine at once with other O_2 molecules to form ozone (O_3). This ozone strongly absorbs the solar radiation over a wide band extending from about 2000 A. to well beyond 3000 A., and the absorption dissociates the O_3 molecules, releasing an O atom again from the O_2 molecule.

Light in the spectral region 1300–1000 A. beyond the λ1760–1350 continuum bands, also shows absorption by O_2. Price and Collins[34], using a source of light giving a continuous spectrum in this region, have shown that O_2 has many strong absorption bands there, their order of intensity being the same as that in the λ1760–1350 region.* The bands observed by Price and Collins are diffuse, and increase in width when the pressure of the absorbing oxygen is increased. These bands have as their limit a wave-length that corresponds to the first ionization potential of O_2 (12·2 V.); however, the efficiency of ionization of O_2 by absorption at this limit seems to be rather low.

The region 1000–670 A. shows further strong oxygen absorption bands, named after Hopfield. These are the strongest of all the O_2 bands, and the light in this region is absorbed completely in a few hundredths of a mm. of O_2 at N.T.P.[34] Hence sunlight of these wave-lengths will be intercepted in the highest layers of the atmosphere. The absorption of such light brings a normal O_2 molecule into one or other of a series of different electronic states, each associated with one or other

* Several observers who have examined this region only with the aid of monochromatic or line sources of light have stated that the oxygen is transparent to light in this spectral region. For example, the first hydrogen line (α) lies in a region in which oxygen is relatively transparent; according to Hopfield 10 cm. of O_2 at N.T.P. is needed to absorb it.

of a series of different amounts of vibrational energy. The limits of the electronic series correspond to ionization. Absorption ejects an electron and also leaves the resulting ion in an excited state. The energies necessary to produce these excited states with their minimum vibrational energy are 16·1, 16·9 and 18·2 e.v., corresponding to convergence limits of the Hopfield bands, of wave-lengths about 770, 730 and 680 A. At the long-wave end of the Hopfield bands, between 950 and 1000 A., it is possible that pre-ionization occurs, that is, absorption by a normal O_2 molecule produces an excited but neutral O_2 molecule, with sufficient energy to emit an electron spontaneously with a small amount of kinetic energy (thus producing an O_2^+ ion as the result of a double process, namely absorption and spontaneous ionization, instead of by a single process, as in the case of the absorption at the limits of the series).

The O_2 absorption in the λ1760–1350 continuum region involves an amount of energy in the solar spectrum which exceeds that in the Hopfield region, and, as already stated, it will cause the oxygen in the outer layers of the atmosphere to be almost entirely in the atomic form, though it must be accompanied by a very small proportion of molecular oxygen. Thus at great heights we shall have a very rare molecular oxygen atmosphere co-existing with much denser atmospheres of N_2 and O. The molecular oxygen atmosphere will absorb light in both the λ1760–1350 region and the Hopfield regions, and the latter absorption will produce ionized oxygen molecules. These may be expected to be rare constituents of the outer atmosphere, much rarer, probably, than N_2^+ ions.

Molecular oxygen ions themselves have absorption bands in the visible and infra-red region (5300–7900 A.), and if O_2^+ were present in sufficient amount this absorption should be evident in the solar spectrum; actually it appears not to be observable, nor would this be expected if O_2^+ is only a very rare constituent of the upper atmosphere, even by day. Even the N_2^+ molecules, which are likely to be present in much greater amount by day, seem insufficient to produce appreciable absorption in the negative-band region of the solar spectrum. Further, though excited O_2^+ will be present in small amounts by day, it will lose its excitation quickly after sunset, so that the emission bands of O_2^+ (5300–7900 A.) are not to be expected—and are not observed—in the spectrum of the night sky. The Runge emission bands of neutral O_2 (4350–3125 A.) also do not appear in the spectrum either of the night sky or of the aurora, though O_2 is likely to be more abundant than O_2^+; their absence may be attributed to a combination of the following causes: (*a*) rarity of O_2 in the high levels where alone any excited O_2 molecules could have time to lose their energy of excitation by radiation, before losing it by collision with another molecule or atom, and (*b*) rarity of excitation of such O_2 molecules as exist at those levels.

§ 13. ATMOSPHERIC PROCESSES INVOLVING OXYGEN ATOMS

The preceding discussion indicates that where nitrogen exists in the upper atmosphere, it will be almost entirely in the molecular state, a small proportion being ionized and excited by day, and reverting to the ionized unexcited state soon after sunset; whereas at high levels the oxygen will exist almost entirely in the atomic state.

These oxygen atoms, when first formed by dissociation, will be partly normal (^3P) and partly excited (^1D), but the latter rapidly lose their energy of excitation by radiation or collision. The O atoms in the outer atmosphere are almost permanently free, since the process of recombination into molecular oxygen is very slow. This process occurs more and more frequently as we descend in the atmosphere, so that we pass from a region of atomic oxygen, through a transition layer in which the proportion of molecular oxygen increases, till we come to a region where atomic oxygen is rare. The level of transition is not yet well determined. At these lower levels such oxygen atoms as are formed by dissociation will rapidly form ozone by combination with O_2; the distribution of the ozone is now known with considerable definiteness through the accordant determinations, by quite different methods, of Götz, Meetham and Dobson[35], and of Regener[36].

Though at high levels the recombination of O to form O_2 is small, the process is very important in enabling the solar energy stored up in the outer atmosphere in the form of energy of dissociation to be transformed into energy of excitation of O atoms, and also of N_2 molecules, during the night. This continuous transformation of energy is probably the source of the light emitted by the atmosphere throughout the night. Each act of recombination of two O atoms into one O_2 molecule, which will usually occur in the near presence of a third body, can impart the energy of dissociation of O_2 (about 5·09 V.) to one or more of the particles involved; this is insufficient to ionize either O or N_2, so that no new ions can be formed in such a process, but it can excite O or N_2 to states from which emission of radiation is possible. Some of these states will be metastable, and a proportion of the O and N_2 in these states may receive further excitation by acting as the third body in other acts of oxygen recombination.

The night-sky spectrum is attributable almost* entirely to neutral molecular nitrogen and atomic oxygen; the latter is the cause, as was shown by McLennan[37] and his collaborators, of the famous yellow-green auroral line λ 5577, and of certain red lines, while the neutral molecular nitrogen produces several bands, notably those bearing the name Vegard-Kaplan. In the auroral spectrum, however, the negative bands due to N_2^+ are prominent, and are naturally explained as due to the simultaneous ionization and excitation of N_2 molecules by the impact of corpuscles from without, which also excite O atoms by impact, to the metastable ^1S state, and possibly also to other "excited" states: the green auroral line is produced by the transition ^1S to ^1D.

§ 14. THE IONIZATION OF ATOMIC OXYGEN

The processes that produce excited O atoms during the night will also give a supply of ^1S and ^1D excited O atoms during the day. Thus there will be large numbers of ^3P (i.e. normal) O atoms, and (probably) much smaller numbers of ^1S and ^1D atoms present; these have excitation energies of 4·2 and 2·0 e.v. They may gain greater energy, in some cases sufficient for ionization, by absorption of sunlight. It is known from well-established spectroscopic data that normal O atoms

* Whether the qualification "almost" can here be omitted is not yet certain, and depends on the confirmation or otherwise of suggested attributions of a few lines to argon, helium and other gases.

(^3P) can absorb light of wave-length $\lambda < 910$ A. (13·55 e.v.) and be transformed into the ^4S state of singly ionized atomic oxygen (O II); likewise that ^1D atoms can absorb light of wave-length $\lambda < 744$ A. (16·5 e.v.) and be transformed into the ^3P excited state of O II; and ^1S atoms can attain the same state by absorption in the region $\lambda < 858$ (14·4 e.v.). In this region of the spectrum these O atoms will compete with N_2 molecules in absorbing the light, and thus attaining the ionized state; their relative success depends on the proportions in which the two gases are present, and their absorption coefficients.

There exists a possibility that O atoms in the ^1S and ^1D excited metastable states could attain the ionized state by absorption in the region $\lambda > 1000$ A., where the competition of N_2 is insignificant; though O_2 absorbs strongly in the same region, the rarity of O_2 in the highest layers as compared with O (and possibly even as compared with the excited forms of O) may give these atoms an opportunity to contribute notably to the ionization there. Absorption of the wave-lengths $\lambda < 1064$ A. (11·6 e.v.) for ^1D atoms, and $\lambda < 1315$ A. (9·4 e.v.) for ^1S atoms, would suffice to raise these O atoms to the ^4S state of ionized atomic oxygen. For instance, absorption might transform ^1S and ^1D atoms into the ^1P state of neutral oxygen (O I), which lies very slightly (6085 cm.$^{-1}$ or 0·75 e.v.) above the ^4S state of O II; in a collision of the second kind, an O I atom in this ^1P state could lose an electron and also a small amount of energy, forming an ionized (O II) atom in the ^4S state. It is worth noting that the wave-length of the light quantum required to raise ^1S atoms to the ^1P excited state is 1217·6 A., which coincides fairly closely with the hydrogen L (α) line, for which $\lambda = 1215{\cdot}7$. Saha has suggested that the ultra-violet end of the sun's spectrum will contain strong emission lines, including L (α); if these exist, they will probably have a width of at least 10 A., and may therefore be potent in ionizing O atoms in the ^1S state.

The ^1D atoms present by day may themselves pass, by the "allowed" transition corresponding to absorption of light of wave-length $\lambda < 999$ A., to the ^1S state, from which further absorption may lead to ionization.

Theoretical calculations by Milne, Pannekoek and others[38] indicate that the probabilities of photo-ionization of atoms are high, at least as high as those for molecules. It can therefore be confidently expected that the photo-ionization processes in the atomic oxygen would cause complete absorption of the corresponding radiation (if the atomic oxygen were absorbing this radiation) in a path-length equivalent to less than 0·01 mm. of O at N.T.P. (containing about 3.10^{16} atoms per sq. cm. section). With a few exceptions (e.g. He and Na and Hg at high temperatures) it has not been possible to study absorption by free atoms in the laboratory; but Collins and Price have observed the ionization-continuum in the spectrum of neutral helium, beginning at 505 A., and R. W. Wood[90] has observed the photo-ionization of sodium atoms. In both these cases extremely small concentrations of the atoms suffice for the appearance of the ionization-continuum in the absorption spectrum. Moreover vacuum spectroscopists working on absorption spectra find that if a small number of free atoms generated in the discharge tube happen to be in the path of the light examined, then strong absorption lines of these atoms occur; for example, L (α) and L (β) always appear in absorption spectrograms, since a few

hydrogen atoms are always present; likewise the oxygen triplet near 1040 Å. appears if any oxygen atoms are present in the discharge tube which is generating the continuous radiation under examination.

§ 15. POLYATOMIC MOLECULES IN THE ATMOSPHERE

The constituent gases of the atmosphere include the rare gases, which are monatomic, the principal gases nitrogen and oxygen which are diatomic, and certain polyatomic molecules, in particular water vapour, carbon dioxide and ozone. All these show their presence in various ways, among which is the formation by them of absorption bands in the solar spectrum. By the study of the Fraunhofer spectrum in the near infra-red, Slipher[41] and his collaborators have shown that no other polyatomic gases are present in the atmosphere in amounts comparable with these; formaldehyde[42] has also been detected. It is desirable, however, that search should be made for absorption bands due to other polyatomic constituents, possibly present in small amounts, such as nitrous oxide (N_2O). Success in this search depends not only on the absolute amounts of these gases present, but also on the situation of their infra-red bands relative to those of the more abundant polyatomic constituents.

Ozone as an atmospheric constituent has so much interest and importance as to have become the subject of an extensive specialized literature; but in view of the recent publication by the Royal Meteorological Society of a symposium of papers on ozone, this subject will not be discussed in detail here. Observation and theory agree in indicating that its percentage amount decreases rapidly above 35 km. In the lower stratosphere O_2 molecules are abundant but O atoms are rare; their combination to form O_3 usually requires the conjunction of a third body, and such are also abundant there. In high levels (probably everywhere above 100 km.) O atoms are relatively abundant, but O_2 molecules are rare, and the necessary third bodies are also much more scarce than farther down. Hence little ozone will be formed there; but N_2 is much more abundant than O_2, and the formation of N_2O, which equally requires the conjunction of a third body, may well be much more probable than that of O_3. The total amount of N_2O thus formed may be low, and yet that amount may be of great importance in determining the temperature of the upper atmosphere. This temperature, on which theoretical calculations have been made by Gowan[43], depends partly on the amount of energy, absorbed from the sun's spectrum, which is converted into kinetic energy of translation (and rotation) of the molecules and atoms, and partly on the rate of "thermal" radiation of such energy. This radiation depends mainly on the polyatomic molecules present. Water vapour, which is lighter than nitrogen, may be present, but this is uncertain, because it absorbs radiation in various ultra-violet regions, including a range of continuous absorption 1785–1550 Å.[44]; this absorption causes dissociation, probably into a normal hydrogen atom and an excited OH molecule (often also with considerable energy divided as kinetic energy between these two particles); the H atom may escape from the atmosphere, and the OH molecule may also become

dissociated by further absorption, in which case water vapour will be continuously lost to the upper atmosphere.*

Carbon dioxide is a relatively heavy constituent of the atmosphere, and its proportion above 100 km. may be distinctly less than in the troposphere. It has an ultra-violet absorption spectrum starting at 1692 A.[44], beyond which there appears a mixture of discrete and diffuse bands which eventually merge into an ionization continuum at 920 A. Other regions of absorption and photo-ionization occur at shorter wave-lengths, but have not yet been systematically explored. The influence of these absorptions on the presence and reactions of gaseous carbon compounds in the upper atmosphere is as yet quite uncertain.

Ozone has the well-known absorption band in the region 2300–3300 A.[45], and also the Chappuis bands in the visible spectrum, as well as infra-red absorption bands. The two former dissociate it, and on this account little ozone can exist at high levels by day, though a little might form during the night.

Nitrous oxide, if formed at all, would probably also be dissociated in the daytime, but may be more abundant than ozone, at high levels, during the night. It may thus have importance in determining the temperature of the outer atmosphere, the atmosphere being heated by day by dissociation.

§ 16. NEGATIVE OXYGEN IONS

Atomic oxygen is known to have a considerable affinity for electrons, the energy of attachment being about 3·3 e.v.[46] Hence O^- ions are likely to occur in the ionized layers; it is uncertain whether the principal process which forms them is combination in the presence of a third body, or a two-body attachment accompanied by radiation. It is not improbable that these ions readily lose their excess electron again by absorption of radiation of quite long wave-length, in which case one effect of the whole process is to prolong the period during which free electrons can avoid their final reunion with one ion of the opposite sign.

The presence of such O^- ions will also enable certain reactions to occur by means of two body collisions which otherwise would require the intervention of a third body. Such reactions are

$$O_2 + O^- = O_3 + e,$$

as suggested by Martyn and Pulley[47] to support the view that ozone is a less rare constituent of the regions above 100 km. than, for example, had been supposed by Chapman; another such reaction, which may be more probable in those regions, because of the greater abundance of N_2 as compared with O_2, is

$$N_2 + O^- = N_2O + e.$$

In both these cases, however, it has to be remembered that O^- ions are probably much rarer than O atoms, so that it is not possible without further examination to

* Water vapour also has a number of discrete bands in the region $\lambda < 1400$, which eventually fuse into an ionization continuum just below 1000 A. (at 12·6 e.v.). Absorption regions from 860 to 740 A. and 694 to 504 A. also probably give rise to photo-ionization, though little is known about the spectra in these regions. These processes probably have little importance for the earth's atmosphere.

say whether such two-body reactions as the above, or the corresponding three-body reactions between N_2 or O_2 and O, are the more probable.

One difficulty in a theory of N_2O equilibrium in the upper atmosphere is that N_2O seems likely, on dissociation, to give rise to NO+O, and NO is not very susceptible to dissociation by absorption: it may, however, combine with oxygen to form NO_2, which will again suffer photo-dissociation to NO[48]; thus we seem to have an irreversible process leading to the formation of NO, which is not known, on present evidence, to be a constituent of the upper atmosphere.

There is some evidence that the photo-dissociation of N_2O does not lead irreversibly to the production of NO. Flory[32] contends that molecular oxygen and molecular nitrogen are the products. A recent paper by A. B. F. Duncan[51] on the far ultra-violet absorption of N_2O indicates that the photo-ionization of this molecule occurs at 12·66 e.v.

REFERENCES

No detailed bibliography of the three following branches of upper-atmospheric physics will be given here; the reader is referred to the recent publication of three summarizing papers or collections of papers, namely:

"Atmospheric ozone" (a symposium): *Quart. J. R. met. Soc.* **62**, Suppt., 1936.

"Ionospheric research" (a symposium): *Proc. Nat. Inst. Sci.*, India, **1**, 1935 (including papers by K. K. Mitra, M. N. Saha and others).

"The Light of the Night Sky": Déjardin, *Rev. mod. Phys.* **8**, 1 (1936).

The *Reports* (1926, 1929, 1932, London, and 1936, Firenze) on *Solar and Terrestrial Relationships*, published by a Committee of the International Research Council (now the International Council of Scientific Unions) also contain many summarizing articles bearing on the upper atmosphere.

(1) See *Meteorological Glossary*, p. 159 (H.M. Stationery Office, 1930).

(2) Regener. *Phys. Z.* **34**, 820 (1933); see also refs. (19), (36).

(3) See Wenstrom. *Mon. Weath. Rev.* (July 1934), and *Proc. Inst. Radio Engrs*, N.Y. (Nov. 1935), for an account of various types of radio transmission devices for atmospheric observation.

(4) See *Ergebn. Kosmischen Phys.* **1**, 1–86 (1931), for a summary by Störmer of knowledge of aurorae up to that date.

(5) Olivier, C. P. *Meteors*, 1925.
Lindemann and Dobson. *Proc. roy. Soc.* A, **102**, 411 (1922).
Öpik, E., *Proc. Nat. Acad. Sci.*, Wash., **22**, 525 (1936).

(6) Störmer. *Geofys. Publ.* **9**, No. 4 (1932).
Störmer. *Avh. Norsk. Vid.-Akad., Mat.-Naturv. Kl.*, No. 2, 1933.

(7) Vestine. *J. R. Astr. Soc. Can.* (1934).

(8) See Whipple. *Quart. J. R. met. Soc.* **61**, 285 (1935), and references there quoted.

(9) Synge. *Phil. Mag.* **9**, 1014 (1930).
Tuve, Johnson and Wulf. *Terr. Mag.* **40**, 452 (1935).

(10) Pedersen. *The Propagation of Radio Waves* (Copenhagen, 1927).
Chapman. *The Earth's Magnetism* (Methuen, 1936).

(11) Pekeris. *Nature*, Lond., **138**, 642 (1936), and references there quoted.

(12) *Proc. roy. Soc.* A, **151**, 105 (1935).

(13) See Paneth. *Roy. Coll. Sci. J.* **6**, 120 (1936).

(14) Hann. *Met. Z.* **20**, 122 (1903), and also Jeans, Humphreys, Wegener (see Jeans, *Dynamical Theory of Gases*, Ch. 15, 1925).

(15) JEANS. *Dynamical Theory of Gases*, Ch. 15 (1925).
(16) ANGENHEISTER. *Terr. Mag.* **37**, 431 (1932).
(17) PANETH and GLÜCKAUF. *Nature*, Lond., **136**, 717 (1935).
(18) DINES. *Quart. J. R. met. Soc.* **62**, 379 (1936).
(19) REGENER. *Nature*, Lond., **138**, 544 (1936).
(20) CHAPMAN and MILNE. *Quart. J. R. met. Soc.* **46**, 357 (1920).
(21) EROPKIN. *Ibid.* **62**, Suppt., 68 (1936).
(22) FUCHS. *Beitr. Geophys.* **47**, 1 (1936).
(23) APPLETON. *Nature*, Lond., **136**, 52 (1935).
(24) VEGARD and TØNSBERG. *Geofys. Publ.* **11**, No. 2 (1935).
(25) MARIS. *Terr. Mag.* **33**, 233 (1928); **34**, 45 (1929).
MARIS and HULBURT. *Phys. Rev.* **33**, 412 (1929).
(26) HARANG. *Terr. Mag.* **41**, 141 (1936).
(27) BIRGE and HOPFIELD. *Astrophys. J.* **68**, 257 (1928).
(28) HOPFIELD. *Phys. Rev.* **36**, 789 (1930).
MULLIKEN. *Phys. Rev.* **46**, 144 (1934).
(29) SLIPHER. *Mon. Not. R. astr. Soc.* **93**, 657 (1933).
(30) HERZBERG. *Naturwissenschaften*, **20**, 577 (1932).
(31) KNAUSS and BALLARD. *Phys. Rev.* **48**, 796 (1935).
(32) FLORY. *J. chem. Phys.* **4**, 23 (1936).
(33) LADENBURG, VAN VOORHIS and BOYCE. *Phys. Rev.* **40**, 1018 (1932).
(34) PRICE and COLLINS. *Phys. Rev.* **48**, 714 (1935).
(35) GÖTZ, MEETHAM and DOBSON. *Proc. roy. Soc.* A, **145**, 416 (1934).
(36) REGENER. *Phys. Z.* **35**, 788 (1934).
(37) MCLENNAN. *Proc. roy. Soc.* A, **108**, 501 (1925); **120**, 327 (1928).
(38) MILNE. *Phil. Mag.* **47**, 209 (1924); *Mon. Not. R. astr. Soc.* **85**, 750, 768 (1925).
PANNEKOEK. *Proc. K. Akad. Wet. Amst.* **29**, 1165 (1926).
CHAPMAN. *Proc. roy. Soc.* A, **132**, (1931).
(39) *Mon. Not. R. astr. Soc.* **89**, 456 (1929).
(40) WOOD. *Physical Optics*, p. 145 (Macmillan, 1935).
(41) ADEL, SLIPHER and BARKER. *Phys. Rev.* **47**, 580 (1934).
ADEL, SLIPHER and FOUTS. *Phys. Rev.* **49**, 288 (1936).
(42) DHAR and RAM. *Nature*, Lond., 25 Nov. 1933.
(43) GOWAN. *Proc. roy. Soc.* A, **120**, 655 (1928); **128**, 531 (1930), and *Quart. J. R. met. Soc.* (Nov. 1936).
(44) RATHENAU. *Z. Phys.* **87**, 32 (1933).
(45) FABRY and BUISSON. *Mem. Sci. Phys.* **11**, 54 (1930).
GÖTZ. *Ergbn. kosm. Phys.* **1**, 180 (1931).
(46) MILLIKEN, *J. Chem. Phys.*, **2**, 782 (1934).
(47) MARTYN and PULLEY. *Proc. roy. Soc.* A, **154**, 455 (1936).
(48) See SPONER. *Molekulspektren*, **1**, 96–128 (1935).
(49) APPLETON, *Quart. J. R. met. Soc.* **62** Suppt., 22 (1936).
(50) *Mem. R. inct. Soc.*, **3**, No. 26 (1930).
(51) DUNCAN. *J. Chem. Phys.* **4**, 638 (1936).

ATOMIC PHYSICS

By N. FEATHER, Ph.D.

§ 1. INTRODUCTION

Experimental results in nuclear physics published since the second report in the present series[1] was written (July 1935) can hardly be said to have contributed any new phenomenon of fundamental importance to the general body of knowledge in this subject. In spite of this, in the year now under review, there has been no slackening in the output of papers describing work in this field. Further and more exact data have been collected in respect of a large number of phenomena, and theorists have had the opportunity, almost for the first time in this domain, of putting various hypotheses to a quantitative test. In this way discrepancies have been revealed which have suggested the need for a far-reaching revision of viewpoint in some cases. In particular they have given rise to a new outlook on the problem of intranuclear forces. The suggestions of Bohr in this matter will be noticed in due course.

Otherwise, this report, the third dealing with the physics of the atom, will follow closely the plan adopted in the second report.

§ 2. ISOTOPIC CONSTITUTION

The position regarding the stable atomic species was given in last year's report[2] roughly as follows: four elements remained to which, at that time and despite frequent attempts,* the methods of mass analysis had not successfully been applied—and eight mass numbers (less than 210) remained which had not already been assigned to any atomic species then known. In a footnote to the main report, Dempster's success with three† of these four elements was briefly recorded, and this, together with the latest research of Aston, was shown to reduce the unassigned mass numbers to four (5, 8, 191, 193). Before the report was in print Dempster[3] also announced the successful analysis of iridium. Two isotopes were found, and it was immediately clear that mass numbers 191 and 193 should be assigned to them. In order to reverse earlier failures and obtain results such as these, the chief requirement was the development of more efficient sources of positive ions for the elements concerned. Dempster investigated several possibilities and finally adopted the method of the high frequency spark[4]. Metal electrodes were employed, separated by a small (adjustable) distance in a high vacuum, and between these a discharge was maintained by the inductive coupling of the spark circuit to the high frequency circuit of a Tesla coil energized from the A.C. mains. With electrodes of a great many pure metals and alloys it was found that a copious supply of metallic ions was

* No attempt had been made with masurium ($Z=43$); this element thus made the fifth for which data were completely lacking.

† Palladium, platinum and gold.

thus produced. Mass analysis of these ions was effected by an arrangement of electric and magnetic fields[5] essentially of the same type as that described by Mattauch and Herzog[6]. These authors may be regarded as having developed the standard apparatus of Aston, by replacing the uniform electrostatic field of the original mass spectrograph by a non-uniform (cylindrical-radial) field. By the use of radial fields not only velocity focusing, but also a considerable degree of geometrical focusing is possible, with a corresponding improvement in the sharpness of "spectral lines" and the consequent possibility of the narrowing of slits for equal times of exposure. In Dempster's apparatus the mean deviations of the ion beam are as follows: by the electrostatic field, 90°; by the magnetic field, 180°.

Following up his initial success with the noble metals, Dempster has recently investigated certain other elements by this method. His results with iron[7], nickel[7], strontium[8], cadmium[9], tellurium[8], barium[10], cerium[10], lead[9] and uranium[11] are as follows:

Fe	54	56	57	*58*				
Ni	58	60	61	62	64			
Sr	*84*	86	87	88				
Cd	106	108	110	111	112	113	114	116
Te	120*	122	123	124	125	126	128	130
Ba	130*	132*	*134*	135	136	137	138	
Ce	136*	138*	140	142				
Pb	204	206	207	208				
U	235*	238						

With four elements asterisks denote mass numbers of isotopes which have not hitherto been detected by the method of mass analysis; italics (3 cases) indicate species just previously[12] discovered and now confirmed. The new results with nickel, cadmium and lead, on the other hand, show the earlier analyses to have been in error in suggesting isotopes which in fact cannot be present with anything approaching the presumed abundance: the isotopes in question are Ni^{56}, Cd^{115} and $Pb^{203, 205, 209, 210}$. Again, the confirmation of $Ni^{61, 64}$ settles a doubt raised by recent conflicting reports of Aston[13] and de Gier and Zeeman[14].

Dempster's negative results with cadmium and lead and those of Bainbridge and Jordan[15] are in entire agreement; the latter authors, moreover, have similarly disproved earlier views concerning the existence of Sn^{121} and Hg^{197}—these species, also, if stable, must be considerably less abundant than has hitherto been supposed. Aston[16], reviewing this new work, has suggested that two additional isotopes of neodymium, $Nd^{148, 150}$, are indicated by a complicated rare-earth mass spectrogram published by Dempster[5]. If this is so, neodymium consists of at least seven isotopic species, the mass numbers previously accepted for this element being 142 to 146, inclusive.

Much important information concerning isotopic constitution has, however, been obtained with instruments of considerably smaller dispersion and resolving

power than those of Dempster and Bainbridge and Jordan.* Thus, de Gier and Zeeman, having brought the old parabola method of Thomson to a high state of perfection, have recently[17] reported an eighth isotope of molybdenum, Mo^{102}, discovered with such an instrument. Their first discovery[18], also, (of A^{38}) has now been entirely confirmed by Nier[19]. According to this last investigation the abundance ratio $A^{38}:A^{40}$ is 1:1660. Nier has been responsible for carrying on and extending the work of Vaughan, Williams and Tate[20] who detected the known rare isotopes of carbon, nitrogen, neon and argon (A^{36})—and determined relative abundances—by the method of mass analysis. They found

$$C^{13}:C^{12}=1:92 \qquad Ne^{21}:Ne^{20}=1:340$$
$$N^{15}:N^{14}=1:265 \qquad A^{36}:A^{40}=1:305.$$

Independently of one another, Nier[21] and Brewer[22] have obtained evidence for a third isotope of potassium (K^{40}).† Their figures for the abundance ratio $K^{40}:K^{39}$ agree remarkably well; we may probably adopt the mean value 1:8500 with confidence. Quite recently Brewer has also reported rare isotopes of lithium (Li^6)[26] and sodium (Na^{22})[27]. The crude abundance ratios here are

$$Li^6:Li^7=1:20{,}000 \qquad Na^{22}:Na^{23}=1:5000.$$

Whilst the existence of the former species is not entirely surprising, the occurrence of the latter in ordinary sodium presents some difficulties. It breaks a well-known regularity in the distribution of mass numbers in the odd-numbered elements[28]—although this difficulty would disappear if F^{18} were discovered also—and, what is more important, it runs contrary to the assumption that Na^{22} is positron active and of comparatively short period[29]. Clearly, further investigation is required.‡

Last year brief reference was made[32] to the method of isotopic hyperfine structure as applied to the investigation of the constitution of the elements. This method, which is more often available for elements of high atomic number than for the lighter elements, in actual fact provided the first information concerning the isotopic complexity of platinum and iridium[33]. Yet, alone, it suffers from the great disadvantage that in any given case only the number of isotopes may be deduced with complete certainty. Mass numbers can never be assigned directly—although, by means of relative abundance estimates and a knowledge of the chemical atomic weight, generally in this matter, also, a fair guess may be made.§ At the present time, however, all the information obtained by optical spectroscopy has been checked and extended by the more fundamental method of mass analysis.

* Determinations of exact masses by these authors are described in § 3.

† The existence of this presumably unstable species had been suspected for some time[23], current explanations of the radioactivity of potassium and rubidium requiring both K^{40} and Rb^{88}. So far, however, the latter species has not been detected in mass analysis: cf. Nier[24], Brewer[25].

‡ Similar cases of present disagreement over the stability of atomic species are those of A^{38} and U^{235}. As already mentioned, mass analysis shows A^{38} to be stable; on the other hand Libby, Peterson and Latimer[30] have suggested that α-activity of short period should be ascribed to this species (see § 4). Again, U^{235} is generally regarded as the ultimate parent of the actinium family; Meitner and Hahn, however, suppose that it is β active and of half-value period 24 min.[31]. It seems more likely in each case that faulty interpretation is somewhere involved than that any true nuclear isomerism has been set in evidence. For the latest results with Na and Li, see reference[107].

§ It is usually possible to decide, when several isotopes are indicated, whether these have consecutive mass numbers or whether certain intermediate species are "missing".

It is merely interesting to note that whilst the first success of the former (hyperfine structure) method was the discovery of Pb^{204} in 1932[34], recently the same method has again been employed with this element[35] to confirm the negative results of Dempster and of Bainbridge and Jordan for Pb^{209}, which have been mentioned above.

During 1935, also, the importance of the band spectrum method was further demonstrated: Aston[13] confirmed the existence of the lighter isotope of indium (In^{113}), first discovered by Wehrli[36] through the analysis of the band spectrum of indium iodide (InI).

§ 3. EXACT MASSES AND SPIN MOMENTS OF NUCLEI

The position with regard to "exact" atomic masses, at any rate for the lighter elements, has once more changed somewhat during the last year[37]. When last year's report was written, disintegration experiments had suggested quite clearly that accepted masses ought to be revised, but only the first beginnings of revision had at that time been made. It is probably correct to say that at the time of writing greater confidence was placed in masses deduced entirely from disintegration data (assuming the validity of the mass-energy relation[93]) than in values obtained in any other way. Now the situation is rather different: in respect of several species, of course, disintegration data still provide the only reliable information, but at present these data are generally[38] used to supplement rather than to replace direct determinations of mass. New determinations by Aston[39] are largely responsible for this change in viewpoint; his particular results are marked by an asterisk in the table which follows. The remaining data in table 1, as before, have been calculated* from determinations of the energy released in various nuclear transformations. It

Table 1

Species	Atomic mass	Species	Atomic mass	Species	Atomic mass	Species	Atomic mass
n^1	1·0090	Be^9	9·0149	O^{17}	17·0046	Mg^{25}	25·9909
H^1	$1{\cdot}0081_1$*	B^{10}	10·0161*	F^{19}	19·0045*	Al^{27}	26·9909*
H^2	$2{\cdot}0147_1$*	B^{11}	11·0128	Ne^{20}	19·9986*	Si^{28}	27·9860*
H^3	3·0170	C^{12}	12·0035*	Ne^{21}	21·0007	Si^{29}	28·9864*
He^3	3·0170	C^{13}	13·0073	Ne^{22}	21·9987	Si^{30}	29·9844
He^4	4·0039*	N^{14}	14·0073*	Na^{23}	22·9971	P^{31}	30·9844
Li^6	6·0168	N^{15}	15·0048	Mg^{24}	23·9938	A^{40}	39·9754*
Li^7	7·0179	O^{16}	16·0000	Mg^{25}	24·9946		

may be said at once that the two sets of data are in general very concordant; at present the chief discrepancies occur in respect of disintegrations in which B^{10} is involved. It would appear that the energy released in the transformations

$$ {}_5B^{10} + {}_0n^1 \rightarrow {}_3Li^7 + {}_2He^4 \qquad \text{.....(1)} $$

and

$$ {}_5B^{10} + {}_2He^4 \rightarrow {}_6C^{13} + {}_1H^1 \qquad \text{.....(2)} $$

is in each case roughly 10^6 electron volts less than would be expected from the masses which are given. Temporarily, however, a discrepancy in this sense is not so serious as one in the other would be (observed energy greater than that pre-

* Cf. Feather, *Introduction to Nuclear Physics* (1936), chap. 5.

dicted), for the missing energy may still be found in the form of radiation, if the direct disintegration transition to the ground state of the final nucleus happens to be a very rare occurrence in these transformations. Alternatively—and from other considerations*—less doubt is thrown on accepted masses by present information concerning reaction (1) than by the data for (2); here, it may be, some change in accepted values is still required. In fact there is already a suggestion of the direction from which such a change may come: Bainbridge and Jordan[41] have re-determined a number of mass differences with a new mass spectrograph—and the only significant disagreement between their results and those of Aston, employed in table 1, is in the case of the doublet ($C^{13}H_4^1 - O^{16}$). The suggestion is that the masses of each of the carbon isotopes may be somewhat greater than is at present supposed. As already indicated, this is more in line with current information regarding disintegration (2).

For the heavier species exact masses are still almost entirely lacking—if "exact" be taken to imply a knowledge of nuclear mass at least to within 10^{-3} mass unit (about 10^6 electron volts energy equivalent)—but no doubt information on this point will soon begin to accumulate. At first it is likely to come mainly from transformation experiments, and, more particularly, from those carried out with slow neutrons. Clearly, a knowledge of the quantum energies of the γ rays emitted in the process of radiative capture in such cases (cf. § 5) will enable one to calculate directly mass differences between isotopes—and determinations of the disintegration energies characteristic of the unstable (β active) species frequently produced in these transformations will serve a similar purpose when isobaric pairs are in question. Meanwhile, it may be noted, a direct attempt to check the validity of the mass-energy relation for the natural radioactive elements has been entirely successful—within the rather wide limits of error which still characterize the determinations of mass in this case. Dempster has employed his high dispersion mass spectrograph to determine the atomic masses of uranium, thorium and the two lead isotopes which constitute the end-products of the radioactive series which derive from these species, respectively. Subject to a re-determination of the masses of certain isotopes of palladium and rhodium, the new values are[42]

U^{238}	238·088	Pb^{206}	206·00 ± 0·01
Th^{232}	232·070	Pb^{208}	208·00 ± 0·01.

Taking Aston's latest value for the mass of the neutral atom of helium (table 1) and the known amount of energy released in the successive disintegrations, the calculated mass differences are as follows:

$$U^{238} - Pb^{206} = 32{\cdot}087$$
$$Th^{232} - Pb^{208} = 24{\cdot}069.$$

It will be seen how good the agreement actually is in both cases.†

* The energy released in the reaction ${}_3Li^7 + {}_2He^4 \rightarrow {}_5B^{10} + {}_0n^1$ (the reverse of (1)) agrees with that deduced from the masses—it is not equal and opposite to the currently accepted energy release in (1): see also a suggestion of Rotblat[40].

† The mass equivalents of the disintegration energies have been taken as $55{\cdot}9 \times 10^{-3}$ and $45{\cdot}8 \times 10^{-3}$ mass units, respectively.

As regards the spin moments of atomic nuclei the methodology of the various types of determination was fairly fully discussed in the last report[43]: it requires now to be supplemented in one respect only. We have to describe that particular development of the method of atomic beams which has made possible the determination of the sign of the nuclear magnetic moment (μ) in certain cases. As was previously indicated[43], so long as a symmetrical deflection pattern is obtained, as in most applications of the method, there is no means of deciding whether the axes of the nuclear magnetic and mechanical moments are similarly or oppositely directed. Rabi[44] has based a method of making this decision on a development of an earlier experiment of Phipps and Stern[45]. At the time of writing only the hydrogen isotopes have been investigated by the new method, but the very important result has thereby been established[46] that, not only for the proton, but for the deuteron also, the positive sign of μ applies; in each case the magnetic and mechanical moments are to be characterized by parallel (as distinct from anti-parallel) axes.

If we are to understand the method by which these results were obtained it will be necessary to discuss first the experiment of Phipps and Stern upon which it was based. The experiment of Phipps and Stern was concerned in some detail with the process of "orientation" of an atom with finite magnetic moment in an applied field: in order to grasp the object of it we shall begin by considering such a magnetic atom devoid of translatory motion in a field of strength H. Pictorially, we should say that, as to rotation, the atom is restricted to those rotations which leave constant the angle between the axis of its resultant magnetic moment and the direction of the field: moreover, only a small number of such angles being permitted, any actual orientation of the atom with respect to the field is one of these permitted orientations, continuously.* Suppose, now, that both the field strength H, and also the direction of the field, are made to vary with time. The problem is to describe the subsequent magnetic behaviour of the atom under discussion. In the investigation of Phipps and Stern the appropriate experimental conditions for the study of this behaviour were realized—by passing a beam of atoms of a given, roughly constant, velocity through a region of space in which the strength and direction of the applied field varied rapidly with the distance traversed. Their results were discussed in terms of the theoretical findings of Güttinger[47] and later with relation to the more complete (if particularized) theory of Majorana[48]. Turning for the moment to this theory, its conclusions were, briefly, that the magnetic behaviour of the atom depends upon the relative values of two angular velocities—the rate of change of direction of the applied field and the "classical" angular velocity of precession of the magnetic atom in the field at any instant. It is to be noted that the spin moments of the nucleus were completely neglected—i.e. assumed to be zero—throughout the calculations. In the first case, when the rate of change of direction of the applied

* An expression of the above restrictions more in terms of observables (and, therefore, one generally to be preferred to a pictorial description) is as follows: we should say that, in the conditions which have been specified, the magnetic potential energy of the atom remains constant (in actual amount one or other of a set of characteristic energies, expressible in terms of H, certain atomic constants and the successive values of a magnetic quantum number, m).

field is small compared with the classical rate of precession, Güttinger and Majorana came to the conclusion (again pictorially expressed) that the atom should adjust itself completely to the changes in the applied field, retaining with respect to the field direction the same orientation throughout. (Expressed in terms of energy changes, the magnetic potential energy of the atom then varies with the field strength, alone, without regard to the changes in the direction of the field—otherwise m remains constant during these changes.) In the other case, when the rate of change of field direction is more rapid (of the same order as the rate of Larmor precession) they concluded, however, that abrupt transitions between different magnetic energy states (sudden changes in m) should be possible: the orientation of the atom need not then follow that of the field completely. Now the experimental results (of Phipps and Stern and of Frisch and Segrè) to which this theory was expected to apply were obtained with potassium: concerning them we need say nothing here except that they were not entirely conclusive—the predictions of the theory did not appear to have been substantiated in all particulars. As already suggested, no further advance was made until Rabi not only pointed out the cause of the apparent discrepancies, but also suggested the important development of the experimental method which is now to be described.

According to Rabi the discrepancies arose from the neglect in the theory to take into account the fact of the finite spin moments of the potassium nucleus. The complicating effects of such moments as regards the behaviour of a magnetic atom in effectively steady fields of different intensities were already understood: as described in last year's report (*loc. cit.*) they formed the basis of the earlier atomic beam experiments of Rabi and his collaborators. In "rapidly rotating" fields these effects emerge somewhat differently. Let us recall, first, the steady field effects. If I and J are the quantum numbers specifying angular momentum for the nucleus and for the extra-nuclear system, respectively, then, in a field of strength H, $(2I+1)\,(2J+1)$ different values of the magnetic potential energy are possible—and we may enumerate these values, which are in general themselves non-linear functions of H, formally as follows:

$2\,(J+I)+1$	different values for atoms with total angular momentum		quantum number	$J+I$
$2\,(J+I-1)+1$	,,	,,	,,	$J+I-1$
.....				
$2\,\lvert J-I\rvert+1$	,,	,,	,,	$\lvert J-I\rvert$.

The difference which the sign of the nuclear magnetic moment introduces at this stage is, briefly, this: the values of the magnetic potential energy which come first in the above enumeration are greater than those which succeed them if the nuclear moment is positive and less than the succeeding values if the nuclear moment is negative. Now in magnetic deflection experiments in effectively steady inhomogeneous fields, if W is the magnetic potential energy in any case, obviously $-\partial W/\partial H$ is the important atomic quantity* determining the deflection. From what

* Deflecting force $= -\partial W/\partial x = -\partial W/\partial H.(\partial H/\partial x)$, and $-\partial W/\partial H$ effectively represents the component of the magnetic moment of the atom in the direction of the field.

has already been said, therefore, this will in general be different, in otherwise similar circumstances, for atoms with a given total angular momentum quantum number (F), according as the nuclear magnetic moment is positive or negative. Conversely if, under specified conditions, we select, by means of a suitable slit, a given portion of the symmetrical deflection pattern resulting from an inhomogeneous field deflection experiment, we shall be selecting atoms in different states (described by different total angular momentum quantum numbers) depending upon the sign of the nuclear moment. The new method of Rabi utilizes the effect of a "rapidly rotating" field upon a beam of atoms so selected in order to decide between the two possibilities of sign: the extent of the non-adiabatic transitions which the rotating field can produce is determined by the values of the total angular momentum quantum number appropriate to the atoms under investigation.

Let us take the simplest case of all, that of hydrogen (atoms of mass number unity), and for the sake of argument assume that the magnetic moment of the proton is positive. From other evidence we already have, for the normal atom, $I=\frac{1}{2}$, $J=\frac{1}{2}$. Then, in a steady field, the enumeration of the magnetic states is as follows:

3 states with total angular momentum quantum number 1 ($\frac{1}{2}+\frac{1}{2}$),

1 state with total angular momentum quantum number 0 ($\frac{1}{2}-\frac{1}{2}$).

Moreover, following our assumption regarding the sign of μ, the single state with $F=0$ is the state of lowest energy—for all values of the applied field. Since, actually, the energy of this state decreases continuously as H increases, the corresponding longitudinal component of the effective magnetic moment of the atom $(-\partial W/\partial H)$ is positive for all values of the field.* Consequently, in a weak (inhomogeneous) field deflection pattern, atoms with $F=0$ will be found in the "stronger field half" of the pattern. Since, in our case—with the hydrogen isotope $_1H^1$—the symmetrical pattern contains four lines (4 magnetic states, in all), it is thus known that one or other of two lines (in one half) is due to atoms for which $F=0$. An exact investigation shows that it must be the less deflected of the two lines which arises in this way. Alternatively, if we had assumed initially a negative value for μ, we should have similarly concluded that the inner of the two lines in the "weaker field half" of the pattern was that formed by atoms with $F=0$. Now, as already indicated, $F=0$ is a single state with no possibilities of non-adiabatic transitions in a rotating field.† On the other hand $F=1$ (which applies to all other lines in the deflection pattern under discussion) includes three magnetic states between which such transitions may occur under suitable conditions—and some of these transitions result in a change in sign of the effective magnetic moment of the atom concerned.‡ It is this last fact which is crucial. Suppose we have isolated the atoms forming

* This is always true of the (sub-) state of lowest energy, whatever the values of I and J for the atom in question.

† Normally the non-adiabatic transitions produced by such a field are between different m states having the same F, transitions from one F state to another do not take place as long as the normal energy difference between these states is large compared with the change in the energy of either produced by the field.

‡ With $_1H^1$, two of the magnetic states for which $F=1$ have negative effective moments and one such state has a positive moment, in all fields, if μ is positive—and vice versa.

one of the inner lines of the weak field deflection pattern for hydrogen. Let these atoms now traverse, in turn, a weak "rotating" field and a strong inhomogeneous field and be received upon a screen. If the atoms are those for which $F=0$, no transitions occur in the rotating field and the strong analysing field merely produces a further deflection in the same direction as before. If, on the other hand, $F=1$ for the atoms concerned, transitions occur as already described—and two lines are formed after deflection in the analysing field. Consequently, if in fact we select the atoms forming the inner line in the "stronger field half" of the pattern produced by the first Stern-Gerlach field, we shall conclude that the magnetic moment of the proton is positive or negative depending upon whether we finally obtain one or two traces after employing the two further fields, as just described. Kellogg, Rabi and Zacharias obtained a single trace in the conditions specified; therefore they concluded that the proton moment was positive. Their decision concerning the deuteron moment was made in a precisely similar manner—though naturally as I increases the entire problem becomes somewhat more complicated.*

The whole importance of a knowledge of the magnitude and sign of the magnetic moment of the deuteron lies in the fact that (taken together with similar knowledge in respect of the proton) it provides the most reliable guide which—in the absence of a direct determination—we possess concerning the neutron moment. Following upon their determination of the signs of the hydrogen moments, Kellogg, Rabi and Zacharias[49] proceeded to determine the magnitude of the deuteron moment more accurately. Their latest value is $0{\cdot}85 \pm 0{\cdot}03$ nuclear magnetons.

Amongst other recent determinations of nuclear magnetic moments that of the less abundant isotope of lithium (Li^6) is perhaps the most interesting. Two[50] different experimental methods have yielded the following results:

$$I=1; \qquad \mu=0{\cdot}6.$$

It appears probable now, from direct determinations in three cases, what was previously[51] merely surmised, that the nuclei of the series H^2, Li^6, B^{10} and N^{14} (possibly the only stable nuclei of even mass number and odd atomic number) have the same nuclear spin ($I=1$) and, in the order of increasing atomic weight, regularly decreasing nuclear magnetic moments ($\mu\ (N^{14}) < 0{\cdot}2$).

In this section of the report we have still to consider the question of electric quadripole moments of nuclei. This question has come into prominence almost entirely within the year under review; it called for only the briefest reference in last year's report[52]. It is not so much that the idea is a new one as that the type of experimental evidence which suggests it has changed completely in very recent times. Pictorially described, a nucleus with a finite electric quadripole moment is one in which the distribution of charge is not spherically symmetrical; formally the quadripole moment, B, is specified as the product of the nuclear charge and an area, q, as follows:

$$B = Ze.q = 2\pi \int_0^{r_0} \int_0^{\pi} \rho r^4 \,(3\cos^2\theta - 1)\sin\theta \, dr\, d\theta \qquad \ldots\ldots(3).$$

* For ${}_1H^2$, $I=1$.

In this specification (r, θ) are polar coordinates (with respect to the centre of the nucleus and its axis of spin) of a point in the nucleus, ρ is the density of electric charge in the neighbourhood of this point and r_0 is introduced as an upper limit of integration in the form of an effective nuclear radius. Let us fix our ideas geometrically by considering a nucleus in which the density of charge is constant throughout ($\rho = \rho_0$). Then, clearly, positive and negative values of q belong to nuclei of which the bounding surfaces are, respectively, prolate and oblate spheroids.* From the experimental side, as was mentioned in the last report[52], evidence for finite quadripole moments has most recently been adduced from the failure of the Landé interval rule in respect of the (magnetic) hyperfine structure components in certain cases. The failure of this rule—or, at least, the apparent failure—is, in itself, no new phenomenon[53]. It has previously been noticed—and explained—in a number of spectra. These earlier explanations[54], however, were based exclusively on known peculiarities of the electron configurations of the atoms concerned: as long as they were successful there was no need to postulate otherwise unknown characteristics for the nuclei. The first instance of the insufficiency of an explanation of this sort occurred with europium—and here the failure was obvious chiefly because the magnitude of the discrepancy with the interval rule was markedly different for the hyperfine structure patterns attributed to the isotopic atoms Eu^{151} and Eu^{153}. It was clear that a nuclear difference was primarily responsible for the effect which had been found[55]. Now the Landé interval rule is the result of the fact that the energies of the individual hyperfine structure terms belonging to a given level are expressible in the general form

$$A_0 + A_1 \cos(IJ) \qquad \ldots\ldots(4),$$

with A_0 and A_1 atomic constants and $\cos(IJ)$ the quantum cosine of the angle between the axes of nuclear and electronic angular momenta (*vide supra*). Empirically, it was found, from an analysis of the europium spectra, that the addition of a third term,

$$A_2 \cos^2(IJ) \qquad \ldots\ldots(5),$$

in the energy expression was both necessary and sufficient to correlate all the data—A_2 being suitably chosen in each case. In Casimir's explanation[56] of this result on the basis of a nuclear quadripole moment, A_2 includes the product of B, the magnitude of this moment (3), and a similar parameter referring to the electron configuration of the atom. If, therefore, either distribution of charge—that on the nucleus or throughout the rest of the atom—happens to be spherically symmetrical, the effect disappears. It would seem that Casimir's explanation is generally accepted at the present time, certainly further examples of the failure of the Landé interval rule have now been discovered[57] and confirmed[58], and the empirical expression of the discrepancies by means of an extra term such as (5) has been found to hold for all the spectra examined. Table 2 contains the results of these investigations expressed in terms of the cross-section q, defined in (3). Clearly,

* Remembering the early attempts to explain the anomalous scattering of α particles in hydrogen in terms of disk-like or spheroidal nuclei, it will be recognized that the assumption of electric quadripole moments is no new development.

data such as these may well become very important for nuclear theory in the near future: at present the most noteworthy feature of them is the fact that large positive values of q occur in some cases. From what has already been said concerning the sign of the quadripole moment it is evident that these positive values cannot be explained merely by ascribing the spin moments of the nucleus to the motion (and intrinsic angular momentum) of a single proton (or neutron) in the spherically symmetrical field of the rest of the nucleus: such motion could only result in the oblate spheroidal "effective shape" which is described by a negative q.

Table 2

Nucleus	Cu^{63}	Cu^{65}	As^{75}	Eu^{151}	Eu^{153}	Lu^{175}	Hg^{201}	Bi^{209}
$q \times 10^{24}$; cm.^{2}	$-0{\cdot}1$	$-0{\cdot}1$	$0{\cdot}2$	$1{\cdot}5$	$3{\cdot}2$	$6{\cdot}1$	$0{\cdot}5$	$-0{\cdot}4$

§ 4. NUCLEAR TRANSFORMATION BY FAST PARTICLES: NEW DATA

Whilst there is no outstanding achievement to be reported in this branch of the subject, many aspects of the problem of nuclear transformation by fast particles are now better understood as a result of the detailed investigations carried out during the past year. It is with this gain in insight, rather than with the mere increase in the body of experimentally determined facts,* that the present section will be chiefly concerned: just because of this preoccupation, moreover, there will be no need to retain the distinction which was made in the last report[60]—transformations which result in stable end-products and those which give rise to radioactive species of short life will be discussed together, as required by the course of the argument. This may conveniently start from the point reached last year in describing the preliminary results of Hafstad and Tuve[61] regarding the transformation of certain light elements by protons.

In a broad survey, the experiments of Hafstad and Tuve may be compared with the much earlier investigations of Pose[62]. In each case the method of the transmutation function—the investigation of the dependence of disintegration yield upon the energy of the responsible particle—was employed, and in each case rather unexpected results were obtained, which afterwards gave rise to considerable discussion. The experiments of Pose, which were carried out with α particles, were soon followed by others of a similar nature and aim: it is no exaggeration to say that present ideas concerning (α, p) disintegrations† derive almost entirely from this general series of experiments. After more than six years, similar experiments are still being performed. There appears to be every indication that the application of the method of the transmutation function to the study of proton- and deuteron-produced disintegrations will lead to equally important developments. As regards deuteron-produced disintegrations—disintegrations of the type (d, p)—the first success is to be attributed to Lawrence, McMillan and Thornton[64]: we shall return to the consideration of their results at a later stage.

* A compilation of recent experimental results has been given by Braddick[59].

† Disintegrations in which an α particle is captured and a proton emitted[63].

The results of Hafstad and Tuve with protons have been extended and further described[65] since the last report was written. Initially they were concerned mainly with the yield of the radioactive product in the known simple capture transformation

$$_6C^{12} + _1H^1 \rightarrow _7N^{13} \qquad \text{.....(6).}^{\bullet}$$

In the range of energies investigated it was found that this reaction takes place chiefly as the result of resonance capture of protons of between 400 and 500 electron kilovolts energy. A further study of the form of the "resonance line" in this case suggested a close doublet—a most interesting result. With lithium and fluorine Hafstad and Tuve investigated the yield of γ radiation as a function of the energy of the protons used for the bombardment. Again very clear-cut resonance effects were found. The latest values for the resonance energies and upper limits for the natural half-widths of the resonance lines are as follows:

$$\gamma \text{ rays from } \begin{cases} Li^7 + H^1 : 440 \pm 11 \text{ ekV.} \\ F^{19} + H^1 : 328 \pm 4,\ 892 \pm 12,\ 942 \pm 15 \text{ ekV.} \end{cases}$$

It will be observed that the results with fluorine again suggest a double line in the high energy range. Many important considerations enter into any attempt to interpret these results. By comparison with the results with carbon, the natural suggestion would be that simple capture of protons (with the formation of Be^8 and Ne^{20} as stable end-products) is involved. Possibly this will prove to be the correct explanation, eventually, but several points remain to be settled before all doubt is removed. In particular we have to discuss the results in question in relation to the much more detailed information which we already possess regarding the transformations

$$_3Li^7 + _1H^1 \rightarrow 2_2He^4 \qquad \text{.....(7)}$$

and

$$_9F^{19} + _1H^1 \rightarrow _8O^{16} + _2He^4 \qquad \text{.....(8):}$$

it is necessary to consider the possibility that the γ rays which are observed may follow similar (p, α) transformations in which the residual nucleus, He^4 in (7) and O^{16} in (8), is left in an excited state. Whilst there is as yet no evidence for this in the form of homogeneous groups of α particles of sub-normal energy in either case, this negative result cannot be regarded as entirely conclusive; such data as we possess[66] regarding the energies of the γ rays suggests that many of the α particles which would be required by this interpretation are of very small energy. On the other hand, the transmutation functions for (7) and (8) are known to be entirely different from those which describe the emission of γ rays. Over the complete range of proton energies employed, no trace of a resonance effect has been reported in either case. Instead, the transmutation function for those disintegrations resulting in α particles of the full energy† are roughly of the type to be expected from the original Gamow[68] theory of the penetration of nuclear potential barriers—allowance being made for the relatively small height of the barrier when protons and light nuclei are

• Further confirmation of the correctness of the interpretation provided by (6) was obtained when Hafstad and Tuve detected the capture radiation produced in the process.

† Amongst recent investigations, see that of Herb, Parkinson and Kerst[67].

involved.* However, it may be argued that this difference, in its turn, is far from conclusive evidence against an explanation in terms of excited product nuclei—for we should not expect the transmutation functions for two alternative reactions (such as those resulting in the same final product in two different states of excitation) to be precisely similar unless incident particles having the same angular momentum about the struck nucleus are chiefly responsible in each case. If, in fact, differences in the transmutation functions originate in this way (because particles with different effective impact parameters are responsible), the differences in question may possibly extend to the positions—or even to the actual occurrence or non-occurrence—of the resonance levels. Now it may well be that the normal and excited states of such nuclei as He^4 and O^{16} have not the same angular momentum. Then, different sets of particles being required for the disintegrations, according to this point of view everything favours the occurrence of the differences in transmutation function which have been found. With plausible explanations on the basis of either assumption (simple capture or disintegration with excitation), clearly a very interesting problem awaits solution here.

Interest in the application of the method of the transmutation function to deuteron-produced disintegrations does not centre chiefly in resonance effects. These have been found in a number of cases, for example with platinum[70], but they do not rank in importance with the discovery of a general difference in form of the yield/energy curves for transformations of the types (d, α) and (d, p), respectively.† As already indicated, this discovery was made by Lawrence, McMillan and Thornton, who studied the transformations of sodium, aluminium, silicon and copper under bombardment by deuterons of energy 0·5 to 3·6 million electron volts. A very clear-cut example of the difference was later reported by Henderson[71]. Under deuteron bombardment the heaviest stable isotope of magnesium is transformed according to the alternative modes

$$ {}_{12}Mg^{26} + {}_1H^2 \begin{cases} \nearrow {}_{11}Na^{24} + {}_2He^4 \\ \searrow {}_{13}Mg^{27} + {}_1H^1 \end{cases} \qquad \text{.....(9),} $$

both products being radioactive. Henderson investigated the corresponding yield/energy relations by studying the intensities of the induced activities obtained under standard conditions.‡ In agreement with the earlier results of Lawrence, McMillan and Thornton he found that the rate of increase of yield with energy was much less rapid in the (d, p) disintegration than in the other. The transmutation function for the (d, α) disintegration was again of the Gamow form. In an attempt to explain this difference an interesting suggestion has been made by Oppenheimer and Phillips[72]. These authors assume that (d, p) transformations take place

* A full discussion of this question in respect of transformation (7), above, has recently been given by Ostrofsky, Breit and Johnson[69].

† The transformations of platinum which show the resonance effects are of the former, (d, α), type. Here, of course, it is just as likely that the resonance conditions refer to the emission of the α particle, as that they characterize the entry of the deuteron into the platinum nucleus.

‡ The possibility of separating the two effects arises in the very different lifetimes of the active products.

essentially by capture of a neutron from the incident deuteron, without the actual penetration of the latter particle into the nuclear structure. Such transformations have been referred to as rejection reactions[71]. Under specified conditions the probability of disintegration by capture of the neutron and rejection of the proton in this way clearly depends upon the energy of binding of these two particles in the deuteron. It is favourable to the hypothesis of Oppenheimer and Phillips that the calculations based upon it fit satisfactorily with the yield/energy curves only when a reasonable value is adopted for this energy of binding. It would thus appear that the suggestion of rejection transformations is a profitable one[73].

We have already discussed the recent experimental work with protons on lithium; it is interesting to remark, therefore, that the transformation of boron by protons has also been the subject of an intensive research during the past year. It will be remembered that these two cases of disintegration were in fact the first to be established using artificially accelerated particles[74]. Moreover, at a very early stage, it was also established that the two disintegration processes were different in a notable respect. Examined under very similar experimental conditions, the disintegration particles from lithium were shown to be monoenergetic, those from boron to possess energies distributed over a wide range. It was suggested[75] that three particles (in this case all α particles) were liberated in the transformation

$$ {}_5B^{11} + {}_1H^1 \rightarrow 3\,{}_2He^4 \qquad \text{.....(10).} $$

If this interpretation were correct it became an interesting problem to examine the distribution of energy amongst the equal products of transformation. Experimentally,* two lines of attack were open—an accurate determination of the distribution in range of the particles ejected from a thin film of boron bombarded by protons of a given energy, and a cloud chamber study of a sufficient number of individual processes, in respect of the ranges and relative directions of emission of the particles. Experiments of the former type were not long in revealing rather unexpected features in the emission; it is only quite recently, however, that the crucial test provided by the latter type of investigation has been successfully applied. At first[77] a great many experimental difficulties were encountered; we shall return to the successful experiments later.

The first unexpected feature in the range distribution curves was discovered by Kirchner and Neuert[78]. These authors found a group of monoenergetic α particles of small intensity just beyond the end of the main distribution. They suggested that it was to be explained in terms of the reaction

$$ {}_5B^{11} + {}_1H^1 \rightarrow {}_4Be^8 + {}_2He^4 \qquad \text{.....(11).} $$

Later, the occurrence of this group of particles was confirmed by Oliphant, Kempton and Rutherford[79] and a second unexpected feature, in the form of a pronounced peak to the continuous distribution of particles, was set in evidence. A general discussion of the results showed, firstly, that the correct explanation of the feeble group of α particles was almost certainly that provided by (11) and, secondly, that, if (10) were taken as the basis of explanation for the remainder of the particles, then

* At an early stage a theoretical discussion was given by Fokker, Belinfante and Kloosterman[76].

the symmetrical mode of disintegration with the three particles emitted in directions making roughly 120° with one another was not the most probable mode. It may justly be said that the expansion chamber experiments of Dee and Gilbert[80] gave to these conclusions the full weight of direct confirmation. By placing a thin target (of pyrex glass) actually within the expansion volume[81], and paying particular attention to the geometrical arrangement of target and cameras, these authors were able to make some very important observations. They have sought to explain their angular distribution, and the range distribution curve of Oliphant, Kempton and Rutherford, in essentially the following terms. The disintegration of boron by protons is primarily a two-particle (p, α) disintegration, like that of lithium, with all contrary appearances to be attributed entirely to the small binding energy of the residual nucleus (Be^8). As in most disintegrations, after emission of the α particle this nucleus may be left in the normal state or excited—and again, as quite frequently happens, the latter alternative is the more probable. The excited state, however, (energy of excitation about 3 million electron volts) is definitely unstable as against particle disintegration. In a time, short compared with that required for emission of radiation, the excited nucleus disintegrates into two α particles. The features of the continuous energy distribution are then explained as follows: the pronounced peak to the curve represents the roughly monoenergetic group of α particles which are ejected opposite to excited beryllium nuclei; the remainder of the curve refers to the pairs of particles liberated when these nuclei disintegrate. The continuous distribution of observed energies here comes from the random distribution of directions of projection of these particles in relation to the initial directions of recoil of the excited nuclei.

Very recently Williams and Wells[82] have applied the method of the transmutation function differentially to the high energy group of monoenergetic α particles and the much more numerous α particles presumed to arise in the two-stage process above described. They found marked resonance effects with the former α particles only, the resonance energy for the impinging protons being 180 ekV. It may be, once again, that an explanation in terms of a spin difference between normal and excited states of the residual nucleus (Be^8) will be suggested here. At any rate, a difference in nuclear spin would reduce the probability of a radiative transition from the higher state to the lower and thus afford a greater opportunity for the break-up of the excited nucleus into two α particles.

We have now discussed in some detail the proton-produced disintegration of boron which gives rise to three α particles. It may be remarked that five other transformations have also been investigated in which three particles are produced: the species Li^7 and B^{10}, when bombarded by deuterons, and B^{10}, C^{12} and N^{14}, when bombarded by neutrons, have been observed to disintegrate in this way. It is natural to enquire whether, in these cases also, the liberation of three disintegration products takes place in two stages which can be separately distinguished. The possibility of a two-stage transformation depends upon the production of an intermediate nucleus in an unstable excited state. In all five cases, formally at least, the production of Be^8 (excited) as this intermediate product would appear to be

possible. Moreover, groups of monoenergetic particles—neutrons and α particles, respectively—have been reported[83] in two of them and ascribed to the transformations

$$_3Li^7 + _1H^2 \rightarrow _4Be^8 + _0n^1 \quad \ldots\ldots(12)$$

and

$$_5B^{10} + _1H^2 \rightarrow _4Be^8 + _2He^4 \quad \ldots\ldots(13),$$

with Be^8 in the ground state in each case. On the other hand, the continuous energy distributions appropriate to the more frequent three-particle processes

$$_3Li^7 + _1H^2 \rightarrow 2_2He^4 + _0n^1 \quad \ldots\ldots(14)$$

and

$$_5B^{10} + _1H^2 \rightarrow 3_2He^4 \quad \ldots\ldots(15)$$

provide conflicting evidence. Whilst no sign of a quasi-homogeneous group of neutrons (corresponding to the production of Be^8 (excited)) is to be found in the former distribution, the expected group of α particles is clearly evident in the latter. Despite the occurrence of reaction (12), therefore, it seems unlikely that the three-particle disintegration which results from the bombardment of lithium with deuterons takes place in two stages as might be imagined. So far, the three-particle disintegrations of B^{10}, C^{12} and N^{14} under neutron bombardment have not been sufficiently studied for a decision on this point to be possible, one way or the other. The three-particle disintegration of nitrogen,

$$_7N^{14} + _0n^1 \rightarrow _3Li^7 + 2_2He^4 \quad \ldots\ldots(16),$$

was recently observed by Bonner and Brubaker[84]. These authors have developed the method of the "artificial" production of intense sources of neutrons for use with the expansion chamber to a high stage of efficiency[85]. The neutrons are generally produced by bombarding targets of lithium or beryllium with deuterons of about 0·8 million electron volts energy produced in a discharge tube. When recoil protons are to be investigated a high pressure expansion chamber containing methane is employed[86]. With a reasonable expenditure of material the tracks of 60,000 protons were photographed in such a chamber in a recent investigation of the distribution of energy amongst the neutrons from (12) and (14)[87]. Similar investigations[88] in respect of the transformations

$$_4Be^9 + _1H^2 \rightarrow _5B^{10} + _0n^1 \quad \ldots\ldots(17)$$

$$_1H^2 + _1H^2 \rightarrow _2He^3 + _0n^1 \quad \ldots\ldots(18)$$

have been carried out on an almost equally lavish scale.* The extent of their data with nitrogen allowed Bonner and Brubaker to establish conclusively the occurrence of a second two-particle disintegration which had often been suspected. This transformation process,

$$_7N^{14} + _0n^1 \rightarrow _6C^{14} + _1H^1 \quad \ldots\ldots(19)^{[89]}$$

they discovered took place with "slow" neutrons as well as with neutrons of high energy: as we shall see later it is most probable that the transformation product, C^{14}, is unstable (β active).

It will be convenient at this stage to discuss some of the more interesting unstable species which have been produced artificially during the period under

* Naturally only a selection from the tracks was submitted to measurement.

review. In such a survey pride of place must almost certainly be assigned to the work of Livingood[90] with bismuth. From a thorough study of the activities obtained under deuteron bombardment, Livingood came to the conclusion that, with particles of sufficiently high energy, the following transformation takes place:

$$_{83}Bi^{209} + {}_1H^2 \rightarrow {}_{83}Bi^{210} + {}_1H^1 \qquad \ldots\ldots(20).$$

This result is not in itself particularly striking until it is remembered that the final product in this case has the same mass and charge numbers as the naturally occurring radioelement, radium E. Now, in actual fact, Livingood found that the artificial product possessed all the properties of this much-studied substance. He followed the growth of artificially produced polonium,

$$Ra\,E \xrightarrow{\beta} Ra\,F \xrightarrow{\alpha} \qquad \ldots\ldots(21),$$

and measured the range of the α particles subsequently emitted. Again, there was complete agreement with previously known data. This result is, surely, sufficiently striking. It appears that an unstable species, belonging to a natural radioactive series, may be produced—and afterwards decay in precisely the normal manner—otherwise than through the spontaneous transformation of its immediate parent in the series.*

From elements of the highest atomic number we pass to the other extreme in reporting what is almost certainly an unstable isotope of helium. Bjerge[91] has discovered an activity of half value period about 0·8 sec. produced by bombarding beryllium with neutrons. The most probable primary transformation is

$$_4Be^9 + {}_0n^1 \rightarrow {}_2He^6 + {}_2He^4 \qquad \ldots\ldots(22).$$

Other light nuclei which are also unstable have been investigated in detail by Fowler, Delsasso and Lauritsen[92] and by many others. Their results are discussed, in connexion with the evidence which they provide for the validity of the conservation laws, elsewhere in this volume[93]. Two species are interesting chiefly for their long lifetime—and the small energy of the electrons which they emit. They have been studied by McMillan[94]. The unstable species in question are Be^{10} and C^{14}—both previously known[79, 80], but not until after this investigation known to be radioactive. Even now, the data must be regarded as bye-products, rather than as results of a single series of experiments designed to provide them: in order to establish artificial radioactivity of slow decay it is necessary to subject the specimen to bombardment over correspondingly long periods of time. This can hardly be economic as a primary experiment. On the other hand, in dismantling any "high voltage" bombardment apparatus which has been in use for a long time, the opportunity for looking for activities of long periods naturally presents itself. McMillan was dealing with the metal parts of a cyclotron chamber which had been in use for about a year. He has estimated the half value period of Be^{10} as at least

* In this connexion a passage from a Trinity College dissertation submitted by the present writer in 1929 has, at least, historical interest: "Obviously . . . we must assume . . . that each atom of radium B existing today upon the earth is the direct descendant of an atom of uranium which suffered disintegration, probably within recent geological times. But the question arises, has the corresponding statement always been true of each atom of radium B which, during the course of, say, solar evolution, ever existed? . . . " (p. 215).

10 years and that of C^{14} as about 3 months. It is probable that in his experiments the primary reactions were

$$ {}_4Be^9 + {}_1H^2 \rightarrow {}_4Be^{10} + {}_1H^1 \qquad \text{.....(23)} $$

and

$$ {}_6C^{13} + {}_1H^2 \rightarrow {}_6C^{14} + {}_1H^1 \qquad \text{.....(24).} $$

We may conclude our survey with two further results. The first concerns an observation of Libby, Peterson and Latimer[30]. These authors irradiated various compounds containing chlorine with the neutrons from a radon-beryllium source, subsequently collecting the chlorine by precipitation as silver chloride. They found that, when this precipitate was heated in a current of gas, a heavy particle activity decaying to half value in about 2 min. was carried away. This activity was studied in a tube counter and also in a low-pressure expansion chamber. In the latter, tracks resembling those due to α particles were obtained having a reduced air range of 3 mm. Since the active gas was not condensed by passage through a copper tube cooled in liquid air it was presumed to be argon. On the other hand, the most likely transformation scheme,

$$ {}_{17}Cl^{37} + {}_0n^1 \rightarrow {}_{17}Cl^{38} \qquad \text{.....(25),} $$

$$ {}_{17}Cl^{38} \xrightarrow{\beta} {}_{18}A^{38} \xrightarrow{\alpha} {}_{16}S^{34} \qquad \text{.....(26),} $$

runs contrary to the general belief in the stability of A^{36}[18, 19] (or of A^{38}, if similar equations be written in respect of Cl^{35}). Moreover, it is hardly likely that an excited state of this nucleus is involved, if a half-value period of 2 min. has to be assigned to it.

The last result of all raises a very interesting problem indeed. Van Voorhis[95] has studied the artificial radioactivity induced in copper under deuteron bombardment. With deuterons of 6 million electron volts energy two different activities were distinguished. Here we are concerned exclusively with the activity of longer life ($12{\cdot}8 \pm 0{\cdot}1$ hours). Van Voorhis found that positrons as well as electrons are emitted by the body possessing this period. He has suggested the scheme

$$ {}_{29}Cu^{64} \begin{cases} \nearrow {}_{30}Zn^{64} + e \\ \searrow {}_{28}Ni^{64} + e \end{cases} \qquad \text{.....(27),} $$

with a branching ratio of about 3 : 2 in favour of electron emission.

§5. SLOW NEUTRON EFFECTS AND THE NUCLEAR MODEL OF BOHR

Effects produced by slow neutrons were mentioned—but only incidentally—in the last report. Now, very little more than a year later, so many and such detailed data are available that a very inadequate treatment is again all that can be attempted. In the nuclear domain probably more attention has been given to investigations of these effects, during 1935, than to any other single branch of the subject. Very soon the question of slow neutron effects will merit a separate report.

Transformations produced by particles of small energy are possible only when two conditions are fulfilled: the transformation process must be exothermic and the projectile must be capable of penetrating the nuclear structure in spite of its

lack of energy. On account of the second condition it is only with neutrons that such transformations can be expected to occur; because of the former, slow neutron-produced transformations are restricted to simple capture processes (with most elements) and, with the lighter elements only, certain (n, α) and (n, p) disintegrations. Since the latter disintegrations have been mentioned already[96], here we shall be concerned entirely with examples of the radiative capture of slow neutrons, and related problems.

In respect of simple capture transformations the initial discovery was due to Fermi[97]. It was found that a number of irregularities in previous work on artificially produced radioactive substances could be attributed to the fact that activation had sometimes been carried out in the vicinity of large amounts of water or paraffin wax. Since multiple scattering of neutrons in hydrogenous substances must rapidly reduce initial velocities to roughly "thermal" values[98], it was realized that a marked variation in capture efficiency with velocity might be in question. Direct experiment quickly confirmed this suspicion: as the velocity of the neutrons was reduced,* in the cases investigated, radiative capture rapidly became very much more frequent. Then it was found that the increase in efficiency varied in a most haphazard way from one element to another. This is a very pronounced—and a most important—feature of the phenomenon. Under standard geometrical conditions an increase of 50 per cent or of a hundredfold might be observed, depending upon the substance bombarded.

Soon this evidence, which was being collected chiefly by observations made upon the artificial radioactivity of the products of the various transformations, was supplemented by exact investigations of the absorption and scattering of slow neutrons[101] (using the disintegration of lithium and boron for purposes of their detection), and of the capture radiation[102] emitted in the primary transformations.† The general aim of many of these investigations has been to deduce precisely the variation of capture efficiency with neutron energy for each element. The results, though altogether incomplete in this respect, already clearly indicate that, even within the range of thermal energies and somewhat above, the variation in capture efficiency is frequently not monotonic. Fairly sharp "absorption bands" occur in the energy spectrum. This fact, the apparent independence of the whole phenomenon of atomic number, and the result[103] that high probability of radiative capture and small probability of elastic scattering often characterize the interaction of neutrons with a single element, may be regarded as the important conclusions to be drawn from these experiments. They must form the basis of the present theoretical discussion.

It may be said at once that the last-mentioned result, alone, is decisive against the earlier attempts at interpretation. Of these the calculations of Bethe[104] may

* Bombardment by slow neutrons is now always achieved by surrounding target and source (producing neutrons of high energy) with water, paraffin wax or other hydrogenous material. The process of slowing down neutrons by this method has been studied by Bjerge and Westcott[99] and Westcott and Niewodniczánski[100], amongst many others.

† These methods, of course, provide evidence of radiative capture which leads to the formation of stable, as well as of unstable, products of transformation. The strong absorption of slow neutrons in cadmium, for example, belongs to the former category.

be taken as representative; they predict a variation in the cross-section for elastic scattering of neutrons which is altogether parallel to that of the cross-section for radiative capture. Two suggestions to remedy this defect in the theory are at present in the field. Bohr[105] has taken as starting point the complete abandonment of the one-particle approximation in the discussion of nuclear phenomena, and Breit and Wigner[106], whilst attempting a treatment which follows more closely the regular methods of wave-mechanics, are also forced, in all considerations of energy levels, to refer these to the system (neutron + nucleus) as a single unit. In this respect the two standpoints have much in common. Bohr's theory, in its particular application to slow neutron effects, makes use of the fact that the nuclear aggregate (neutron + nucleus), before the radiation of energy, will generally have energy in excess of the normal by nearly 10^7 electron volts. We start, then, by enquiring what will be the nature of the stationary states of this system for approximately this energy of excitation. A decision here obviously depends upon the assumptions which are made concerning the way in which excess energy is stored in the nucleus. If it be supposed that additional energy is taken up by some quantized collective motion of all the nuclear particles together, then clearly the greater the amount of energy so absorbed the greater also the multiplicity of nuclear states in which it is capable of being retained. Nuclear energy levels rapidly become more and more closely spaced the greater the energy of excitation. From this point of view there need be no contradiction between the γ ray data, on the one hand—exhibiting widely spaced energy levels in heavy nuclei for 1 or 2 million electron volts energy of excitation—and the fact, demonstrated by the occasional occurrence of resonance effects in slow neutron capture, on the other, that at the higher excitations (say of 10^7 electron volts excess energy) the probability that an energy level shall be found in any arbitrarily chosen small interval of energy (say 0–10 electron volts for the incident neutron) is not excessively remote. The haphazard incidence amongst the elements of greatly enhanced capture at low velocities, and the evidence of sharp absorption bands, now become mutually explanatory facts.

The new model, also, is capable of explaining the relatively small extent of the elastic scattering of slow neutrons in strongly absorbing materials. The argument is roughly as follows. Scattering is regarded as normally a dual process; a neutron enters the nuclear structure and, after a finite time, a neutron is again emitted. Clearly such re-emission can only take place when sufficient energy is once more concentrated in a single nuclear particle; when the incident neutron is of small energy this may take a very long time. In such circumstances the emission of radiation, resulting in permanent capture, is a much more probable occurrence. This is a most important point: for the time being, at least, it has rendered the ideas of Bohr central to any theoretical discussion of the interaction of slow neutrons with matter. Whether the formal development of these ideas, when it has finally been completed, will exhibit them as equally fruitful where high energy charged projectiles are in question is indeed an interesting speculation. For, with such particles, there appears at present to be much less necessity for regarding disintegration phenomena energetically otherwise than as single events.

REFERENCES

(1) FEATHER. *Rep. Prog. Phys.* **2**, 74 (1935).
(2) *Report* (1934), § 2.
(3) DEMPSTER. *Nature*, Lond., **136**, 909 (1935).
(4) DEMPSTER. *Nature*, Lond., **135**, 542 (1935).
DEMPSTER. *Rev. sci. Instrum.* **7**, 46 (1936).
(5) DEMPSTER. *Proc. Amer. phil. Soc.* **75**, 755 (1935).
(6) HERZOG. *Z. Phys.* **89**, 447 (1934).
MATTAUCH and HERZOG. *Z. Phys.* **89**, 786 (1934).
See also HENNEBERG. *Ann. Phys.*, Lpz., **19**, 335; **20**, 1 (1934).
(7) DEMPSTER. *Phys. Rev.* **50**, 98 (1936).
(8) DEMPSTER. *Phys. Rev.* **50**, 186 (1936).
(9) DEMPSTER. *Proc. Amer. phil. Soc.* **75**, 755 (1935).
(10) DEMPSTER. *Phys. Rev.* **49**, 947 (1936).
(11) DEMPSTER. *Nature*, Lond., **136**, 180 (1935).
(12) DE GIER and ZEEMAN. *Proc. K. Akad. Wet. Amst.* **38**, 959 (1935).
BLEWETT and SAMPSON. *Phys. Rev.* **49**, 778 (1936).
(13) ASTON. *Proc. roy. Soc.* **149**, 396 (1935).
(14) DE GIER and ZEEMAN. *Proc. K. Akad. Wet. Amst.* **38**, 810 (1935).
(15) BAINBRIDGE and JORDAN. *Phys. Rev.* **49**, 416 (1936).
(16) ASTON. *Nature*, Lond., **137**, 613 (1936).
(17) DE GIER and ZEEMAN. *Proc. K. Akad. Wet. Amst.* **39**, 327 (1936).
(18) DE GIER and ZEEMAN. *Proc. K. Akad. Wet. Amst.* **37**, 127 (1934).
(19) NIER. *Phys. Rev.* **49**, 272 (1936).
(20) VAUGHAN, WILLIAMS and TATE. *Phys. Rev.* **46**, 327 (1934).
(21) NIER. *Phys. Rev.* **48**, 283 (1935).
(22) BREWER. *Phys. Rev.* **48**, 640 (1935).
(23) *Report* (1934), § 2.
(24) NIER. *Phys. Rev.* **49**, 272 (1936).
(25) BREWER. *Phys. Rev.* **49**, 856 (1936).
(26) BREWER. *Phys. Rev.* **49**, 635 (1936).
(27) BREWER. *Phys. Rev.* **49**, 856 (1936).
(28) *Report* (1934), § 2.
(29) FRISCH. *Nature*, Lond., **136**, 220 (1935).
(30) LIBBY, PETERSON and LATIMER. *Phys. Rev.* **48**, 571 (1935).
(31) MEITNER and HAHN. *Naturwissenschaften*, **24**, 158 (1936).
(32) *Report* (1934), § 4.
(33) FUCHS and KOPFERMANN. *Naturwissenschaften*, **23**, 372 (1935).
VENKATESACHAR and SIBAIYA. *Nature*, Lond., **136**, 65, 437 (1935).
(34) SCHÜLER and JONES. *Naturwissenschaften*, **20**, 171 (1932).
(35) ROSE and STRANATHAN. *Phys. Rev.* **49**, 916 (1936).
(36) WEHRLI. *Helv. phys. Acta*, **7**, 611 (1934).
(37) *Report* (1934), § 2.
(38) BONNER and BRUBAKER. *Phys. Rev.* **49**, 19 (1936).
OLIPHANT. *Nature*, Lond., **137**, 396 (1936).
POLLARD and BRASEFIELD. *Nature*, Lond., **137**, 943 (1936).
(39) ASTON. *Nature*, Lond., **137**, 357, 613 (1936).
(40) ROTBLAT. *Nature*, Lond., **138**, 202 (1936).
(41) BAINBRIDGE and JORDAN. *Phys. Rev.* **49**, 883 (1936).
JORDAN and BAINBRIDGE. *Phys. Rev.* **50**, 98 (1936).
(42) DEMPSTER. *Nature*, Lond., **138**, 120, 201 (1936).
(43) *Report* (1934), § 4.
(44) RABI. *Phys. Rev.* **49**, 324 (1936).
(45) PHIPPS and STERN. *Z. Phys.* **73**, 185 (1931).
See also FRISCH and SEGRÈ. *Z. Phys.* **80**, 610 (1933).

(46) KELLOGG, RABI and ZACHARIAS. *Nature*, Lond., **137**, 658 (1936).
(47) GÜTTINGER. *Z. Phys.* **73**, 169 (1931).
(48) MAJORANA. *Nuovo Cim.* **9**, 43 (1932).
(49) KELLOGG, RABI and ZACHARIAS. *Phys. Rev.* **49**, 867 (1936).
(50) FOX and RABI. *Phys. Rev.* **48**, 746 (1935).
SCHÜLER and SCHMIDT. *Z. Phys.* **98**, 430 (1936).
(51) GOLDHABER. *Proc. Camb. phil. Soc.* **30**, 561 (1934).
See also TOLANSKY. *Z. Phys.* **78**, 71 (1932).
(52) *Report* (1934), § 2.
(53) SCHÜLER and JONES. *Z. Phys.* **77**, 801 (1932).
(54) CASIMIR. *Z. Phys.* **77**, 811 (1932).
GOUDSMIT and BACHER. *Phys. Rev.* **43**, 894 (1933).
(55) SCHÜLER and SCHMIDT. *Z. Phys.* **94**, 457 (1935).
(56) CASIMIR. *Physica*, 's Grav., **2**, 719 (1935).
See also SCHÜLER. *Phys. Z.* **36**, 812 (1935).
(57) SCHÜLER and SCHMIDT. *Z. Phys.* **98**, 430; **100**, 113 (1936).
(58) MINTZ and GRANATH. *Phys. Rev.* **49**, 196 (1936).
(59) BRADDICK. *Ann. Rep. Chem. Soc.* **32**, 15 (1936).
(60) *Report* (1934), §§ 5 and 6.
(61) HAFSTAD and TUVE. *Phys. Rev.* **47**, 506 (1935).
(62) POSE. *Phys. Z.* **30**, 780 (1929).
POSE. *Z. Phys.* **64**, 1 (1930).
(63) Cf. BOTHE. *Phys. Z.* **36**, 776 (1935).
(64) LAWRENCE, MCMILLAN and THORNTON. *Phys. Rev.* **48**, 493 (1935).
(65) HAFSTAD and TUVE. *Phys. Rev.* **48**, 306 (1935).
HAFSTAD, HEYDENBURG and TUVE. *Phys. Rev.* **49**, 866 (1936).
(66) CRANE, DELSASSO, FOWLER and LAURITSEN. *Phys. Rev.* **48**, 531 (1934).
CRANE, DELSASSO, FOWLER and LAURITSEN. *Phys. Rev.* **48**, 125 (1935).
(67) HERB, PARKINSON and KERST. *Phys. Rev.* **48**, 118 (1935).
(68) GAMOW. *Z. Phys.* **52**, 510 (1928).
(69) OSTROFSKY, BREIT and JOHNSON. *Phys. Rev.* **49**, 22 (1936).
See also OSTROFSKY, BLEICK and BREIT. *Phys. Rev.* **49**, 352 (1936).
(70) CORK and LAWRENCE. *Phys. Rev.* **49**, 788 (1936).
(71) HENDERSON. *Phys. Rev.* **48**, 855 (1935).
(72) OPPENHEIMER and PHILLIPS. *Phys. Rev.* **48**, 500 (1935).
(73) Cf. SNELL. *Phys. Rev.* **49**, 555 (1936).
(74) COCKCROFT and WALTON. *Proc. roy. Soc.* **137**, 229 (1932).
(75) COCKCROFT and WALTON. *Nature*, Lond., **131**, 23 (1933).
(76) FOKKER, BELINFANTE and KLOOSTERMAN. *Physica*, 's Grav., **1**, 705 (1934).
(77) KIRCHNER. *Naturwissenschaften*, **21**, 473 (1933).
DEE and WALTON. *Proc. roy. Soc.* **141**, 733 (1933).
(78) KIRCHNER and NEUERT. *Phys. Z.* **35**, 292 (1934).
(79) OLIPHANT, KEMPTON and RUTHERFORD. *Proc. roy. Soc.* **150**, 241 (1935).
(80) DEE and GILBERT. *Proc. roy. Soc.* **154**, 279 (1936).
(81) Cf. DEE. *Nature*, Lond., **132**, 818 (1933).
(82) WILLIAMS and WELLS. *Phys. Rev.* **50**, 186 (1936).
(83) BONNER and BRUBAKER. *Phys. Rev.* **48**, 742 (1935).
COCKCROFT and LEWIS. *Proc. roy. Soc.* **154**, 246 (1936).
(84) BONNER and BRUBAKER. *Phys. Rev.* **49**, 223 (1936).
(85) Cf. *Report* (1934), § 5.
(86) BRUBAKER and BONNER. *Rev. sci. Instrum.* **6**, 143 (1935).
(87) BONNER and BRUBAKER. *Phys. Rev.* **48**, 742 (1935).
(88) BONNER and BRUBAKER. *Phys. Rev.* **47**, 910 (1935).
BONNER and BRUBAKER. *Phys. Rev.* **49**, 19 (1936).
(89) BONNER and BRUBAKER. *Phys. Rev.* **49**, 778 (1936).
(90) LIVINGOOD. *Phys. Rev.* **49**, 876 (1936).

(91) BJERGE. *Nature*, Lond., **137**, 865 (1936).
(92) FOWLER, DELSASSO and LAURITSEN. *Phys. Rev.* **49**, 561 (1936).
(93) MAY. This volume, p. 89.
(94) MCMILLAN. *Phys. Rev.* **49**, 875 (1936).
(95) VAN VOORHIS. *Phys. Rev.* **49**, 876 (1936).
(96) *Report* (1934), § 5; also this report, reference (89).
(97) FERMI, AMALDI, PONTECORVO, RASETTI and SEGRÈ. *La Ricerca Scientifica*, **5** (2), 282 (1934).
(98) CONDON and BREIT. *Phys. Rev.* **49**, 229 (1936).
MOON. *Proc. phys. Soc.* **48**, 648 (1936).
(99) BJERGE and WESTCOTT. *Proc. roy. Soc.* **150**, 709 (1935).
(100) WESTCOTT and NIEWODNICZÁNSKI. *Proc. Camb. phil. Soc.* **31**, 617 (1935).
(101) DUNNING, PEGRAM and FINK. *Phys. Rev.* **47**, 325 (1935).
DUNNING, PEGRAM, FINK and MITCHELL. *Phys. Rev.* **48**, 265 (1935).
MITCHELL and MURPHY. *Phys. Rev.* **48**, 653 (1935).
(102) RASETTI. *Z. Phys.* **97**, 64 (1935).
FLEISCHMANN. *Z. Phys.* **97**, 242, 265 (1935).
KIKUCHI, HUSIMI and AOKI. *Nature*, Lond., **137**, 992 (1936).
(103) DUNNING, PEGRAM, FINK and MITCHELL. *Phys. Rev.* **48**, 265 (1935).
(104) BETHE. *Phys. Rev.* **47**, 747 (1935).
(105) BOHR. *Nature*, Lond., **137**, 344 (1936).
(106) BREIT and WIGNER. *Phys. Rev.* **49**, 519 (1936).
(107) SAMPSON and BLEAKNEY. *Phys. Rev.* **50**, 456 (1936).

THE CONSERVATION OF ENERGY AND MOMENTUM IN ELEMENTARY PROCESSES

By A. NUNN MAY, Ph.D., Wheatstone Laboratory,
King's College, London.

§ 1. INTRODUCTION

The mechanical principles of the conservation of energy and momentum are immediate consequences of Newton's laws of motion and therefore lie at the base of all classical mechanics. Their domain has been extended, by postulating additional forms of energy, such as that of the electro-magnetic field, and they have been taken over into the quantum mechanics without any essential modification. These laws were deduced from observations on macroscopic bodies and their extension to atomic dimensions requires independent confirmation.

It is convenient to distinguish two possible ways in which the conservation laws might break down.

Firstly some process may always involve the disappearance (or appearance) of energy or momentum, unaccompanied by any observable change in the rest of the universe. Such a "one-way" non-conservation could in principle be detected by macroscopic experiments, and the methods for the detection of single particles would only be used in this case because of their convenience. We shall discuss this type of non-conservation in connexion with the β-ray disintegration. We may point out that it is generally possible to rescue the conservation laws in cases of this type by postulating new forms of energy.

Secondly the conservation laws may hold if the average of a large number of individual processes is taken, but the individual processes themselves may not obey the conservation laws. This statistical interpretation was proposed by Bohr, Kramers, and Slater for all processes involving radiation. Their theory has been experimentally tested for the Compton effect, and we shall discuss the results obtained under that heading. In order to detect this type of non-conservation it is necessary to observe each of the particles taking part in one individual process. This can only be done in rather special cases, when the particles are charged and have a high velocity.

In the following we shall discuss a number of processes from this point of view. For the most part we shall deal with the interaction of high energy particles, because with these new types of phenomena, involving non-conservation, might be expected to occur. We have classified the experimental evidence according to the type of particles taking part. Thus § 2 is devoted to the mutual interaction of heavy particles (atomic nuclei, neutrons etc.). Phenomena connected with the emission of light particles (electrons, positrons) are discussed in § 3 while the interaction of radiation and matter is dealt with in § 4. In each case a general familiarity with the phenomena is assumed so that attention may be confined to the relevant experimental evidence.

§ 2. COLLISIONS BETWEEN ATOMIC NUCLEI

Elastic collisions; cloud-chamber work. The investigations of Blackett and his collaborators[1] on the forked tracks due to collisions between α-particles and other nuclei in the cloud chamber are well known. Their work is summarized in *Radiations from Radioactive Substances*[2] and more recently Blackett and Lees[1] have given a careful discussion of the accuracy of these measurements. Some measurements have also been made by Wells and White[3] on the collisions of fast protons passing through hydrogen.

It is found in these experiments that the conservation laws apply to the majority of collisions within the experimental error, but there are a few cases of apparent non-conservation. Amongst these the forks corresponding to disintegrations can generally be distinguished by the different appearance of the track of the ejected proton from that of the incident α-particle. There remain *some* forks in which there is no sign of disintegration, but nevertheless the conservation laws are violated, for instance the three branches of the fork may not lie in one plane. [This might be due to the ejection of a neutron without capture of the α-particle, except that such a process involves a large absorption of energy (10 MV.*).] These anomalous forks are generally ascribed to experimental accidents such as the scattering of one of the particles by an atom of the gas, close to the point of collision. However, until recently this explanation had not been established experimentally.

Occasional examples of apparent non-conservation are also observed when photographs are taken of the emission of α-particles by radioactive bodies, such as actinon, in the cloud chamber. If a sufficiently low gas pressure is used, the track of the recoiling Ac A nucleus can be seen. It should of course lie in the same line as the track of the α-particle. Such photographs have been taken by Akiyama[4]. He found that the track of the recoiling Ac A nucleus frequently made angles up to 15° with the track of the α-particle. It was difficult to explain this by a subsequent collision between the Ac A and an atom of the gas, because in such a collision the deflection of the heavy Ac A nucleus could not exceed 4°. Akiyama made the unlikely suggestion that some very hard radiation was emitted at the moment of disintegration.

This phenomenon was investigated in greater detail by Joliot[5]. By using still lower pressures of gas ($\sim$1 cm. of mercury) he was able to distinguish a third track whenever the Ac A recoil was not in the expected direction. This shows quite clearly that the apparently anomalous disintegrations are due to the deflection of the recoiling Ac A nucleus by a collision between it and one of the atoms of the gas. The probability of such a collision was found to be very high. It is reasonable to extend the same explanation to the anomalous forked α-particle tracks.

The fact that the heavy recoil nucleus can be deflected through large angles by collisions with the comparatively light gas atoms has still to be explained. Joliot investigated these collisions, and found many examples of scattering through larger angles than would be expected. He suggests various explanations, none of which is satisfactory. The present position is that anomalous recoil tracks, and therefore

* Throughout this report MV. will stand for million electron volts.

probably the anomalous forks, are accounted for by scattering, but the details of the process are still in doubt.

Disintegrations. The test of Einstein's equation. In discussing the great wealth of experimental work on nuclear disintegration it is only possible to touch on a few problems which are of special interest. Any disintegration may be used to test the conservation of energy as expressed in Einstein's law of equivalence of mass and energy $E=mc^2$.* For this purpose the masses of the nuclei must be known accurately, by means of mass spectrograph work; and the energy released in the disintegration must be known.

We may take as an example the disintegration

$$^{14}N + {}^{2}D \rightarrow {}^{12}C + {}^{4}He.$$

This has been observed by Lawrence, McMillan and Henderson[6], and by Cockcroft and Lewis[7]. The experimental value of the energy release is $13{\cdot}22 \pm 0{\cdot}1$ MV. This is obtained from the range of the α-particles in mica, making allowance for the energy carried away by the recoiling ^{12}C nucleus.†

This might be compared with the difference between the masses on the two sides of equation by substituting the masses given by Aston[8]. However the errors in these masses are not independent. The primary measurements from which the masses are deduced are the doublet separations, e.g. the difference between the values of m/e for the ions $^{1}H_2^+$ and $^{2}D_1^+$. To obtain a reliable estimate of the accuracy the energy changes in disintegrations must be calculated directly from these measurements. This method has been used by Bainbridge in discussing several disintegrations.

We give here the doublet separations obtained by Aston and Bainbridge separately, as there is some discrepancy between them.

We have

	Aston	Bainbridge and Jordan[9]
$^{14}N - {}^{12}C\ {}^{1}H_2 =$	$-0{\cdot}01245 \pm 0{\cdot}00007$	$-0{\cdot}0130 \pm 0{\cdot}0002$
$^{1}H_2 - {}^{2}D =$	$0{\cdot}00152 \pm 0{\cdot}00004$	$0{\cdot}00153 \pm 0{\cdot}00004$
$^{2}D_2 - {}^{4}He =$	$0{\cdot}02551 \pm 0{\cdot}00008$	

By addition

$$(^{14}N + {}^{2}D) - (^{12}C + {}^{4}He) = 0{\cdot}01458 \pm 0{\cdot}00011 \qquad 0{\cdot}01404 \pm 0{\cdot}00022$$

The above are in terms of $^{16}O = 16{\cdot}0000$. Converting to units of energy, by using Einstein's equation, we obtain

Aston	$13{\cdot}6 \pm 0{\cdot}15$ MV.
Bainbridge	$13{\cdot}1 \pm 0{\cdot}2$ MV.
Experiment	$13{\cdot}22 \pm 0{\cdot}1$ MV.

The agreement is probably as good as can be expected at the present stage. The Einstein law is verified within about $\pm 0{\cdot}3/13 = 2$ per cent.

* We shall regard Einstein's law as an accurate expression of the conservation law, and interpret any deviation from it as due to non-conservation. We shall see that within the experimental accuracy conservation holds, and so the question of whether an alternative formulation of Einstein's law would be justified does not arise.

† In correcting for the energy of recoil, conservation of momentum has to be assumed.

The same type of calculation can be carried through for the reaction

$$^6\text{Li} + ^2\text{D} \rightarrow 2\,^4\text{He}.$$

Bainbridge[10] has measured the doublet separation $^6\text{Li} - ^2\text{D}_3$, assuming a deuteron mass of 2·01363, he obtained 6·0145 for Li. Hence his doublet separation, was 0·0264 with an error ± 0·0003.*

$$\therefore \quad ^6\text{Li} + ^2\text{D} - 4\,^2\text{D} = -0{\cdot}0264 \pm 0{\cdot}0003$$
$$4\,^2\text{D} - 2\,^4\text{He} = \quad 0{\cdot}0510 \pm 0{\cdot}00016 \quad \text{(Aston)}$$
$$\therefore \quad ^6\text{Li} + ^2\text{D} - 2\,^4\text{He} = \quad 0{\cdot}0246 \pm 0{\cdot}0003$$

which $= 22{\cdot}88 \pm 0{\cdot}3$ MV.

The experimental value[13] of the energy release is 22·06 ± 0·07 MV. The difference 0·8 ± 0·3 is not significant. Again the Einstein law is verified with an accuracy of 4 per cent.

These two reactions provide perhaps the best verification of the Einstein law because of the large energy releases involved. Many examples of smaller energy releases could be given, for instance

$$^{19}\text{F} + ^1\text{H} \rightarrow ^{16}\text{O} + ^4\text{He},$$

for which Aston's figures give 8·1 ± 0·3 MV., and experiment[11] 8·25 ± 0·2, and finally an example discussed by Jordan and Bainbridge[12]

$$^{14}\text{N} + ^2\text{D} \rightarrow ^{15}\text{N} + ^1\text{H},$$

for which the mass data give 8·57 MV. ± 0·2 and experiment[7] 8·53 ± 0·1 MV.

The accuracy in the mass spectrograph measurements is lower than that attainable in the measurement of disintegration energies. It is possible to devise tests of the conservation laws involving only disintegration data, and these have a higher accuracy. The following three disintegrations have been investigated very thoroughly, and established by experiments with Wilson cloud chambers, and with separated isotopes.

$$^7\text{Li} + ^1\text{H} \rightarrow 2\,^4\text{He} + 17{\cdot}06 \pm 0{\cdot}06 \text{ MV.} \quad \text{.....(1),}$$
$$^6\text{Li} + ^2\text{D} \rightarrow 2\,^4\text{He} + 22{\cdot}06 \pm 0{\cdot}07 \text{ MV.} \quad \text{.....(2),}$$
$$^6\text{Li} + ^2\text{D} \rightarrow ^7\text{Li} + ^1\text{H} + 5{\cdot}0 \pm 0{\cdot}05 \text{ MV.} \quad \text{.....(3).}$$

The values for the energy releases are those given by Oliphant, Kempton, and Rutherford[13]. Subtraction of equation (3) from (2) gives (1). It will be seen that the energy checks within an uncertainty of 100,000 eV. Expressed as a fraction of the total energy release involved, 44 MV., this is 0·2 per cent. However this agreement would still exist if the *same proportion* of the energy of disintegration were lost in each disintegration by such processes as radiation, interaction with neutrinos, outer electrons, or simple non-conservation. It is, of course, unlikely that these processes would take place to exactly the same extent in three different disintegrations, but this point must be remembered in assessing the accuracy of this check on the conservation laws.

* This determination is rather old, and was made by less reliable methods than the others quoted here. This particular example should therefore be treated with reserve.

Until recently the heavy, natural, radioactive elements could not be used for a test of Einstein's equation. The difference between the accepted atomic weights of uranium and radium for instance could be compared with the total mass and energy emitted in the radioactive series between these two elements, and similarly for radium → Ra G and thorium → thorium D.

If this was done discrepancies of the order of 0·2 in the atomic weights or 2×10^8 eV. were found. These were almost certainly due to the inaccuracy of the chemical atomic weight determinations.

Very recently Dempster[14] has made the first accurate determinations of the masses of uranium and thorium, and has followed this by a determination of the masses of the end products of these two series Ra G and Th D. These new masses now agree with the conservation laws, within the rather large limits of error. For instance the total emission of energy in the uranium series is 52 MV. and the mass change gives 57 ± 10 MV.

The conservation of momentum. So far we have not considered the evidence on the conservation of momentum,* nor the question of whether the energy of disintegration is strictly constant or has only a statistical meaning.

We may first consider the evidence on this point obtained in experiments with a cloud chamber. The older work is mainly concerned with the disintegration of nitrogen by α-particles according to the equation

$$^{14}\mathrm{N} + {}^{4}\mathrm{He} \rightarrow {}^{17}\mathrm{O} + {}^{1}\mathrm{H}.$$

Blackett and Lees[1] have collected and discussed the data on the eight forks which are most suitable for exact measurement. With the exception of one anomalous fork (of which the explanation is probably that offered above for the anomalous elastic collisions) these showed a mean energy change ($-1{\cdot}3$ MV.) nearly the same as that found in experiments using electrical counting. The mean deviation from this value (0·4 MV.) was that to be expected from the experimental errors. The forks also showed by their coplanarity that momentum was conserved.

Later work using the cloud chamber on other disintegrations has confirmed these conclusions. Thus, using artificially accelerated ions, Dee and Walton[15] showed that the two α-particles from the disintegration

$$^{7}\mathrm{Li} + {}^{1}\mathrm{H} \rightarrow 2\,{}^{4}\mathrm{He}$$

were emitted in opposite directions with equal momenta, and Dee[16] showed that the same applies in the disintegration

$$^{2}\mathrm{D} + {}^{2}\mathrm{D} \rightarrow {}^{3}\mathrm{H} + {}^{1}\mathrm{H}.$$

Another instance of the same type is the disintegration of boron and lithium by slow neutrons

$$^{6}\mathrm{Li} + {}^{1}n \rightarrow {}^{4}\mathrm{He} + {}^{3}\mathrm{H},$$

$$^{10}\mathrm{B} + {}^{1}n \rightarrow {}^{7}\mathrm{Li} + {}^{4}\mathrm{He}.$$

* Since energy and momentum form one four-vector in relativity theory, the conservation of energy in any process might be regarded as a sufficient guarantee that momentum is also conserved. The conservation of energy is moreover a much more sensitive test of losses by radiation or neutrinos. However for the sake of completeness we consider separately the direct evidence on the conservation of momentum.

Traces caused by these disintegrations in a photographic emulsion show that the two particles are ejected in opposite directions[17]. Since the neutrons causing the disintegration are *slow* there is no allowance to be made for the initial momentum.

In disintegrations by neutrons, Kurie[18] has suggested that the straightforward application of the conservation laws as above may be no longer valid. He investigated the reaction

$$^{14}N + {}^{1}n \rightarrow {}^{11}B + {}^{4}He,$$

and found a constant value, not for the difference between the kinetic energies of the particles on the two sides of the equation, but for the sum of the kinetic energies of the $^{11}B + {}^{4}He$ alone.* He suggested that the neutron was first captured to form a radioactive ^{15}N, emitting its surplus energy as γ-radiation, and that the ^{15}N then split up into ^{11}B and ^{4}He with a definite "proper disintegration energy". This theory was discussed in the last report. Since then Bonner and Brubaker[19] have also obtained evidence of this kind.† There are strong theoretical objections to this novel type of disintegration, and particularly in view of the large errors which can occur in measuring the neutron energy it should be regarded with reserve.

Is the energy of disintegration constant? Other evidence on the conservation of energy in the individual disintegrations can be drawn from the homogeneity of the groups in absorption curves taken with counters. For instance in the disintegration

$$^{7}Li + {}^{1}H \rightarrow 2{}^{4}He,$$

Oliphant, Kempton and Rutherford[13] compared the α-particle group with that from a radioactive source. The disintegration group was found to be less homogeneous than the radioactive one, but only by an amount to be expected from the experimental conditions (thick target etc.). Another such example is the disintegration

$$^{10}B + {}^{2}D \rightarrow {}^{11}B + {}^{1}H,$$

in which Cockcroft and Lewis[7] obtained a proton group of 90 cm. range, which appears to be homogeneous within ± 4 cm. The upper limit to any actual inhomogeneity set by this type of experiment is about $\pm 20{,}000$ eV. This is much lower than the limit set by the cloud chamber experiments.

A more exact limit to the possible inhomogeneity cannot be obtained when the energy is determined by means of the range, because of the straggling. At present more precise methods, such as magnetic analysis, cannot be used because of the small intensities available.

We may however use the essentially similar process of α-disintegration among the heavy radioactive elements. The groups of α-particles emitted by say Ra C′ have been measured with great precision by magnetic analysis[20]. The apparent inhomogeneity is about $\pm 10{,}000$ eV. and is completely accounted for by the geometrical conditions (slit-width). The actual inhomogeneity must therefore be much smaller than this. Any process leading to inhomogeneity, such as radiation, of interaction with the outer electron shells would be expected to be more active amongst the heavy nuclei.

* The velocities being measured with respect to the centre of gravity of the system.
† This has since been withdrawn[56].

§ 3. THE END POINT OF THE CONTINUOUS β-RAY SPECTRUM

We come now to a problem where the breakdown of the conservation laws has been suggested in many quarters, and which has been extensively studied from this point of view.

A full discussion of the evidence on the β-ray spectra will be found in the *Report of the International Conference on Physics*[21] in 1934. Since then a considerable amount of theoretical and experimental work has been done which has made the situation much clearer, but there are still some features which require further study.

It has been known for a long time that the electrons emitted in β-ray disintegrations have a continuous distribution in energy. The experiments of Ellis and Wooster[22], and of Meitner and Orthmann[23] established that the total energy emission, as detected in a calorimeter was given by the *mean* energy of the β-rays. It followed that the β-rays were actually emitted from the nucleus with a continuous distribution in energy. On the other hand there was a good deal of independent evidence that nuclei of the same species were identical. In order to reconcile these two statements Bohr[24] suggested that the conservation laws broke down in the β-ray disintegration (in a macroscopic way, cf. § 1). On the other hand Pauli[25] made the assumption that at each disintegration an uncharged particle, the "neutrino", was emitted which was able to pass out of the calorimeter (in the above experiments) without losing its energy.* The partition of the energy of disintegration between the electron, the neutrino and the residual nucleus, accounted for the continuous spectrum. Attempts to detect any ionization by the neutrino[26] have failed, and the only evidence for its reality is its success in accounting for the facts of the β-ray disintegration.

On the neutrino theory the energy of disintegration should be equal to the *maximum* energy of the β-ray spectrum (neglecting small corrections for the recoil of the nucleus, and the (undetermined) mass of the neutrino). Recent experimental work has mainly been concerned with testing this statement. To do this we must: (1) determine the energy of disintegration indirectly by (say) measuring the masses of the initial and final nuclei, or by using other disintegrations to determine this mass difference; (2) determine the end point of the β-ray spectrum.

This last is especially difficult because the spectrum has no abrupt end, but is found to approach zero intensity very slowly. If a reliable theoretical formula were known for the shape of the β-ray spectrum then this could be used to extrapolate the observed curves, and give an accurate value of the end point. The first theory making use of the neutrino hypothesis was put forward by Fermi[27], but the form of his β-ray spectrum was too symmetrical. The theory was modified by Konopinski and Uhlenbeck[28]† in such a way as to remedy this. Their formula seems to give

* The neutrino was also given a spin $\frac{1}{2}h/2\pi$ in order to obtain conservation of angular momentum in the β-disintegration.

† Konopinski and Uhlenbeck introduced an interaction involving *derivatives* of the wave function of the neutrino, i.e. depending on its *velocity* as well as on its position. This does not seem to be capable of a simple physical interpretation, but is justified empirically by its results.

good agreement with the observed curves at least over the greater part of the distribution. However, extrapolations according to this theory lead to values of the end points which are generally beyond any energies actually observed, and are sometimes difficult to reconcile with indirect determinations of the disintegration energies.

The end point of the continuous spectrum. In the *Conference Reports* the identity of the end point with the disintegration energy was based on experiments on the double disintegration of ThC. The measurements of Henderson[29] on the end points of the β-ray spectrum of ThC and ThC″ showed that if the end point did represent the actual energy release in the disintegration then the total energy released along the two branches

was practically the same, the figures being 11·20 MV. for the upper branch, and 11·19 MV. for the lower. However, if it were assumed that the actual energy releases are given by the mean β-ray energies, the two branches would still agree fairly well, because the β-ray spectra of Th C and Th C″ are not very different. Using the mean energies given by Sargent[30] we obtain 9·75 MV. along the upper branch and 9·98 MV. along the lower, a discrepancy which is not large enough definitely to exclude this hypothesis.

Experiments on the artificial radioactive elements lead to more definite evidence on this point. In some recent papers in the *Physical Review*[31,32] the form of the β-ray (or positron) distribution has been determined for a number of elements. The curvatures of the tracks in a Wilson chamber in a magnetic field were measured. The distribution curves obtained were compared with the Konopinski-Uhlenbeck theory (see above) and give a satisfactory agreement. The end points determined by inspection and by extrapolation were compared with indirect determinations of the disintegration energy. These calculations are given here with changes arising from some recent redeterminations of the quantities involved.

The two radioactive elements ^{8}Li and ^{12}B have spectra extending up to about 10 MV. and therefore provide an excellent test of whether the mean or the maximum of the spectrum is equal to the disintegration energy, since the maximum is about 7 MV. greater than the mean. This advantage is set off by the uncertainty in the interpretation of the experiments.

The Element. ^{8}Li. Crane, Delsasso, Fowler and Lauritsen[33] observed the emission of high energy electrons when lithium targets were bombarded with deuterium ions. They ascribed these to the formation of the unstable element ^{8}Li by the reaction

$$^{7}\mathrm{Li} + {}^{2}\mathrm{D} \rightarrow {}^{8}\mathrm{Li} + {}^{1}\mathrm{H} + Q_1 \qquad \ldots\ldots(1).$$

The half-period was found to be 0·5 sec., the disintegration probably being either

$$^8\text{Li} \rightarrow {}^8\text{Be} + \epsilon^- \qquad \text{.....(2),}$$

or

$$^8\text{Li} \rightarrow 2\,{}^4\text{He} + \epsilon^- \qquad \text{.....(3).}$$

The difference in energy emission between these two would be only slight, since ^{8}Be is known to have nearly the same mass as two α-particles.

In order to obtain the mass of ^{8}Li it is necessary to observe the protons in reaction (1). This is difficult since they are masked by the protons from

$$^6\text{Li} + {}^2\text{D} \rightarrow {}^7\text{Li} + {}^1\text{H} \qquad \text{.....(4).}$$

Delsasso, Fowler and Lauritsen[34] have reported a group of protons of 26 cm. range of about the right intensity, besides the 30 cm. group from reaction (4).* If reaction (1) takes place at all, the protons must have a range less than 30 cm., or they would have been observed in previous experiments. From these considerations we may put Q_1 certainly less than 4·3 MV.

Comparing this with the reaction

$$^7\text{Li} + {}^2\text{D} \rightarrow {}^8\text{Be} + {}^1n + 14{\cdot}3 \text{ MV.} \qquad \text{.....(5).}^{[35]}$$

or

$$^7\text{Li} + {}^1\text{H} \rightarrow 2\,{}^4\text{He} \qquad \text{.....(6),}^{[13]}$$

we find that the energy release in the disintegration of ^{8}Li according to (2) or (3) must be at least 10·2 MV. This figure agrees with the observed upper limit to the spectrum and cannot be reconciled with the *average* β-ray energy which is only about 3·5 MV.

The element ^{12}B. Very similar considerations apply to the other short period element obtained by Crane etc., ^{12}B. This is supposed to be formed by the reaction

$$^{11}\text{B} + {}^2\text{D} \rightarrow {}^{12}\text{B} + {}^1\text{H} \qquad \text{.....(7).}$$

However, a careful search for these protons has been made by Cockcroft and Lewis[7] who could not find any group of sufficient intensity beyond 2 cm. range. Disregarding this difficulty, a similar comparison may be made (assuming a very low energy for the protons) and the result is again, that the energy of disintegration must be the maximum and not the mean energy of the β-rays.

Positron emission ^{13}N. We turn now to some cases of positron emission, to which the same theoretical considerations apply. A much discussed disintegration is that of ^{13}N. Although the positrons have only a comparatively low energy, the quantities involved have been measured accurately, so that a fairly severe test of the conservation principle can be obtained.

The calculations have been made by a number of workers including Cockcroft and Lewis[7], Tuve and Hafstad[36], and Newson[37]. They are given here with the latest available data. The method is to compare the three reactions

$$^{12}\text{C} + {}^2\text{D} \rightarrow {}^{13}\text{N} + {}^1n + Q_8 \qquad \text{.....(8),}$$

$$^{12}\text{C} + {}^2\text{D} \rightarrow {}^{13}\text{C} + {}^1\text{H} + Q_9 \qquad \text{.....(9),}$$

$$^{13}\text{N} \rightarrow {}^{13}\text{C} + \epsilon^+ + Q_{10} \qquad \text{.....(10).}$$

* Recent experiments[57] with separated isotopes have shown that this new group actually belongs to reaction (4) and not to (1). The protons predicted by equations (1) and (7) have not been found despite careful searches. This reduces considerably the value of any argument based on the existence of ^{8}Li and ^{12}B.

Clearly we should have

$$Q_{10} = Q_9 - Q_8 - (n - \mathrm{H}) - (\epsilon^- + \epsilon^+)^* \qquad \ldots\ldots(11).$$

Bonner, Delsasso, Fowler and Lauritsen[38] give $Q_8 = -0{\cdot}37 \pm 0{\cdot}03$ MV.

Cockcroft and Lewis[7] state that $Q_8 > -0{\cdot}32$ MV.

We may adopt the value $Q_8 = -0{\cdot}32$ MV.

For Q_9 there are three accurate determinations.

Cockcroft and Lewis give $2{\cdot}66 \pm 0{\cdot}02$ MV.

Bonner, Delsasso, Fowler and Lauritsen give $2{\cdot}65 \pm 0{\cdot}07$ MV.

Bainbridge and Jordan[9] by an accurate determination of the doublet separation $^{12}\mathrm{C} + {}^1\mathrm{H} \rightarrow - {}^{13}\mathrm{C}$ gives $2{\cdot}76 \pm 0{\cdot}1$ MV. This is important because it excludes the possibility of a high energy γ-ray emission in reaction (9).

The mass of the neutron is best obtained from Feather and Bretscher's[39] work on the photo disintegration of the deuteron,† which together with Aston's mass for the latter gives

$$n - \mathrm{H} = 0{\cdot}83 \pm 0{\cdot}05 \text{ MV}.$$

Combining these values we obtain from equation (11)

$$Q_{10} = 1{\cdot}13 \pm 0{\cdot}15 \text{ MV}.$$

The maximum energy of the positrons from $^{13}\mathrm{N}$ has been determined by many workers; the values obtained are generally somewhat higher than the above. Thus Fowler, Delsasso and Lauritsen[31] give a distribution which appears to end at 1·25 MV., but when extrapolated according to the theory of Konopinski and Uhlenbeck, gives an end point at $1{\cdot}45 \pm 0{\cdot}1$ MV. Kurie, Richardson and Paxton[32] obtained almost the same result. Alichanow, Alichanian and Dzelepow[41] give a curve ending at 1·45 MV., and Cockcroft, Gilbert and Walton[42] one at 1·1 MV.

These values mostly agree with the predicted value within the possible errors, although the discrepancy in the case of the extrapolated end point of 1·45 MV. is large enough to throw some doubt on the Konopinski-Uhlenbeck theory.

The element $^{17}\mathrm{F}$. This element has been formed in two ways which we will consider separately.

The reaction $$^{16}\mathrm{O} + {}^2\mathrm{D} \rightarrow {}^{17}\mathrm{F} + {}^1n + Q_{12} \qquad \ldots\ldots(12),$$

has been studied by Newson[37]. From the excitation curve he obtains

$$Q_{12} > -1{\cdot}8 \text{ MV}.$$

The F_{17} disintegrates according to the equation

$$^{17}\mathrm{F} \rightarrow {}^{17}\mathrm{O}\epsilon^+ + Q_{13} \qquad \ldots\ldots(13),$$

and this may be compared with

$$^{16}\mathrm{O} + {}^2\mathrm{D} \rightarrow {}^{17}\mathrm{O} + {}^1\mathrm{H} + Q_{14} \qquad \ldots\ldots(14).$$

As before $$Q_{13} = Q_{14} - Q_{12} - (n - \mathrm{H}) - (\epsilon^- + \epsilon^+).$$

Cockcroft and Lewis[7] give $Q_{14} = 1{\cdot}91$ MV. which leads to $Q_{13} \leqslant 1{\cdot}9$ MV.

* ϵ^- and ϵ^+ stand for the (equal) masses of the positive and negative electrons. The presence of ϵ^- in equation (11) arises from the unfortunate convention of inserting, not the nuclear masses, but the masses of neutral atoms, in these equations.

† The experiment reported by Ising and Helde[40] leads to $(n - \mathrm{H}) = 0{\cdot}4$ MV. but this is almost certainly incorrect.

The observed distribution curve extends to 2·1 MV. According to the Konopinski-Uhlenbeck theory it ends at 2·4 MV. so that here again too high a value is obtained.

The same active element is formed by the reaction

$$^{14}N + {}^{4}He \rightarrow {}^{17}F + {}^{1}n + Q_{15} \qquad \text{.....(15).}$$

According to Haxel[43] the excitation of neutrons starts at an α-particle energy of 5·9 MV. which leads to $Q_{15} = -4{\cdot}6$ MV.

Using the reaction $^{14}N + {}^{4}He \rightarrow {}^{17}O + {}^{1}H + Q_{16}$ .....(16),

with $Q_{16} = -1{\cdot}3$ MV.

we obtain
$$Q_{15} = Q_{16} - Q_{15} - (n - H) - (\epsilon^- + \epsilon^+)$$
$$= 1{\cdot}5 \text{ MV.}$$

which is in still worse agreement with the experimental value. The deduction of energy changes from excitation curves is not very reliable, but the results can hardly be in error by as much as 1 MV.

The element ^{30}P. The energy balance in the disintegration of ^{30}P has been discussed by Ellis and Henderson[44]. They found that the maximum positron energy was 2·9 MV. This result can be compared with the data for the two reactions

$$^{27}Al + {}^{4}He \rightarrow {}^{30}Si + {}^{1}H + Q_{17} \qquad \text{.....(17)}$$

$$^{27}Al + {}^{4}He \rightarrow {}^{30}P_3 + {}^{1}n + Q_{18} \qquad \text{.....(18),}$$

Q_{17} has been determined as 2·07 MV. by Duncanson and Miller[45]. Q_{18} was determined by Jaeckel[46], from the recoil tracks produced by the neutrons in a cloud chamber. From his measurements $Q_{18} = -2{\cdot}3$ MV. He states that this is likely to be an underestimate but actually all his evidence of high energy neutrons was obtained from protons projected at large angles to the neutron beam, and is therefore unreliable. By comparing equations (17) and (18) we obtain 2·55 MV. as the predicted maximum positron energy. The probable error in this is difficult to estimate but it is at least 0·4 MV., so that the difference between this and the experimental figure is not serious.

The above five elements are the only ones for which data of any accuracy are available, the results are collected in the following table.*

Element	+ or − active	Energy change (MV.) from other disintegrations	Observed end point	K − U end point
^{8}Li	−	> 10·2	10 MV.	11·2 MV.
^{12}B	−	> 12·1	11·5	13·0
^{13}N	+	1·13 ± 0·15	1·1 → 1·4	1·45 ± 0·1
^{17}F	+	< 1·9, < 1·5	2·1	2·4
^{30}P	+	2·55 ± 0·4	2·9 ± 0·1	

* Since this was written Bonner and Brubaker[58] have given similar calculations for the element ^{11}C.

It will be seen that although the energy of disintegration is certainly not given by the mean particle energy, the accuracy with which it can be equated to the maximum is not very great. The end point of the spectrum found by inspection gives a fair agreement with the energy of disintegration as determined indirectly from other disintegrations, and within the experimental error these may be identified. On the other hand the end point obtained by extrapolation on the Konopinski-Uhlenbeck theory appears to be rather too high:* this throws doubt on the method of extrapolation. Thus we may regard the neutrino hypothesis as justified by the agreement between the end point and the disintegration energy, but the form of the spectrum near the end point is not yet given correctly by the theory.

We may mention as a further possible method of confirming the neutrino theory, the experiment carried out by Leipunski[47]. By a very ingenious method he succeeded in measuring the velocity of the atoms ^{11}B, recoiling when radioactive ^{11}C disintegrates. He found that the velocity was considerably greater than would be expected from the ejection of the electron alone, and ascribes the extra velocity to the ejection of the neutrino. The *maximum* velocity of recoil should be no greater with a neutrino than without. In both cases it occurs when the electron is ejected with all the available energy, leaving the neutrino, if it exists, at rest. The presence of the neutrino may however cause an increase in the *average* velocity of recoil. This seems to be indicated by Leipunski's curve. Further work on these lines should lead to more definite results.

§ 4. THE INTERACTION OF RADIATION AND MATTER

The question of whether the conservation of energy and momentum holds in detail, or only statistically for processes involving radiation has been a centre of a controversy during the past year which is a curious repetition of one which took place in 1924. Owing to the difficulty at that time of reconciling the photon theory of light with the electromagnetic theory, Bohr, Kramers, and Slater[48] proposed a less crude application of the particle conceptions to radiation problems.

In terms of this theory one may picture a typical process of radiation as follows. In the disintegration

$$^{10}B + {}^{4}He \rightarrow {}^{13}C + {}^{1}H$$

the ^{13}C nucleus is sometimes left in an excited state. It subsequently reaches the ground state by emitting a γ-ray. The mean life τ of the ^{13}C nucleus in this excited state is extremely short, 10^{-12} sec. Thus within 10^{-12} sec. of the emission of the proton, a quantum will be emitted which may be recorded by a Geiger counter if it gives rise to a secondary electron inside it. This is the particle picture. According to the Bohr-Kramer Slater theory the nucleus is a source of radiation which decays exponentially according to the law $e^{-t/\tau}$. This radiation falls on the counter and

* An agreement between the extrapolated end point of the β-ray spectrum, and the disintegration energy is sometimes said to prove that the mass of the neurino is zero or very small. Dr Peierls has pointed out to the author that this is incorrect. The insertion of a finite neutrino mass in the Konopinski-Uhlenbeck theory modifies the extreme end of the spectrum so that an extrapolation such as that used in these experiments no longer gives the end point of the spectrum, but still gives the true disintegration energy (i.e. including the neutrino mass).

gives rise to a certain probability, that a secondary electron will be ejected and a count recorded. This probability is quite independent of whether any other secondary electron has been ejected by the same source of radiation, or whether the ^{13}C nucleus has yet fallen to its ground state (although this fall is not "observable"). Apart from this the predictions of the Bohr Kramer Slater theory are just the same as the particle theory. In practice τ is always extremely short, at least for phenomena where the individual acts of radiation can be detected, and so within experimental limits on either theory there should be coincidences between the emission of the proton and of the discharge of the Geiger counter. This is found to be so in practice[49].

This agreement disappears when we consider the Compton effect, because here two secondary processes, recoil electron and scattered quantum (detected as photo electron) follow from the same primary process. The "photon" theory of the Compton effect is well known, it leads to certain equations between the angles of projection θ, ϕ of the scattered quantum and the recoil electron, and the energies of these particles, and of course the two secondary processes happen simultaneously.

As applied to the Compton effect the Bohr Kramer-Slater theory on the other hand proposed that a volume element containing a certain density of electrons, and irradiated by the primary beam, had a certain probability per unit time of ejecting a recoil electron in any direction θ. (The electron would have the energy corresponding to this direction.) Also there was a certain probability that somewhere along a direction ϕ a photo electron would be ejected, corresponding to the scattered quantum.

Here also the relation between the energy of the photo-electron and the angle ϕ was left untouched. The theory however regarded these two probabilities as quite independent, it denied that the two events (recoil electron, photo-electron) were simultaneous except by chance, and that the relation given by the particle theory between the angles θ, and ϕ was valid.*

This theory agreed with all the facts which were known at that time. It was tested: (1) by Bothe and Geiger[50] who detected the scattered electron in one Geiger counter, and the photo-electron in another, and found coincidences in time between the discharges of the counters. (There was no test of the relation between θ and ϕ since the scattering took place inside the electron counter.) (2) by Compton and Simon[51] who introduced a beam of X rays into a cloud chamber where the scattering took place. They were able to see the track of the recoil electron, and in some cases also the photo-electron ejected from some lead plates by the scattered quantum. They were therefore able to verify the relation between θ and ϕ. Both

* In the absence of any correspondence between the individual recoil electrons and photo-electrons, the angular conditions necessarily fail. The above description of the theory is correct if the primary beam is considered as a continuous radiation field. Difficulty is sometimes found in seeing that the same result also follows even when the primary radiation is regarded as made up of separate pulses (e.g. from individual electrons arriving at the anti-cathode of the X-ray tube). It is true that both the scattered electron and photo-electron have to coincide with such a pulse, but this does not ensure coincidences of the two. Suppose the counter detecting recoil electrons responds to one pulse in every 10^{18} and the quantum counter to one in 10^{19}. Then if, as on the B.K.S. theory, the counters "choose" their pulses at random and independently, there is little chance of their choice coinciding.

these experiments were of great technical difficulty, but the results were sufficiently definite to disprove the B. K. S. theory.

The first of the recent repetitions of the Bothe-Geiger experiment was that of Shankland,[52] which led to the conclusion that the particle theory was incorrect. As a source of radiation Shankland used a tube containing radon (100 mg.) in equilibrium with its products. After being collimated by a long hole in a block of lead the γ-rays struck the target, for which various materials were used. The recoil electrons and the scattered γ-rays were detected by two systems of counters which were protected from the primary beam and from "stray" scattering by means of lead shields. For the scattered γ-rays five counters were used in line, thus increasing the chance of recording the γ-ray quantum by a factor of five (but also increasing the natural count five times).

The recoil electrons were also detected by tube counters; two used as a coincidence counter in one experiment and one alone in another. These counters were set to detect particles coming at definite angles, generally 35° to the primary beam for the electrons.*

The number of coincidences due to chance was calculated from the observed rates of counting in the individual counters and the resolving time, the latter having been determined previously by finding the number of coincidences when the counters were discharged by a radium source.

In the actual experiment it was found that the observed number of coincidences only slightly exceeded the number due to chance, the difference being much lower than that expected from the particle theory. Further no more coincidences were observed when the counters were placed in positions where the coincidences would be expected, than when one of the counters was rotated out of the plane formed by the primary beam and the other counter, when on the particle theory no coincidences should occur.

Thus with a paraffin scatterer 0·05 cm. thick Shankland obtained 14 ± 3 coincidences per hour, compared with an expected number of 69 on the photon theory, and 9 due to chance. When one of the counters was rotated through 90° about the primary beam, the observed number of coincidences was still 12 ± 4 per hour.

From these results Shankland concluded that the particle theory was incorrect. Before passing to other experiments we may notice that the whole argument depends on the correctness of the predicted number of coincidences. If this were smaller by a factor of ten, it would hardly be detected in these experiments. Other workers using the more homogeneous Th C^{11} γ-rays, have generally found that the actual number of coincidences fell short of that expected from the geometry by a large factor. This discrepancy is principally due to inhomogeneity in the γ-ray beam, and to scattering of the recoil electrons in the target. It is difficult to judge how important these factors were in Shankland's experiments, but it is probably along these lines that an explanation of his results must be sought.

* The angle for which the γ-ray counter was set is not given, it appears to be about 80° on the diagram. This would lead to coincidence for a primary quantum energy of about 500,000 eV. The angular conditions are not sharp and quanta between 250,000 and 800,000 eV. might give coincidences. This illustrates the difficulties caused by using an inhomogeneous source; there is no definite angle at which the counters should be set.

The publication of Shankland's results was followed by many speculations[53] on their interpretation, for instance whether the contradiction between his results and those of Bothe and Geiger was due to the use of γ-rays or to the test of angular conditions as well as simultaneity. This discussion need not be gone into now in view of the results obtained in other experiments.*

Among the later experiments supporting the particle theory, which have been reported briefly, we may take those of Bothe and Maier-Leibnitz[54] for discussion. They used a source of radio thorium (20 mg.), the γ-rays from which are much more homogeneous than those from Ra (B + C). They used a thin scatterer of celophane, and one counter each for the electrons and scattered quanta set so as to detect particles coming at an angle of 30° with the primary beam.

The presence of the scatterer increased the number of counts in the electron counter by a factor of two, from 45 per min. to 85 per min., but made no appreciable difference to the quantum counter (40 per min.) owing to its much lower efficiency. The coincidences numbered 25 in $14\frac{1}{2}$ hours. The number of these due to chance was determined by removing the scatterer and bringing the number of electron counts up to the same number as with the scatterer by means of a radium D source. (This method of determining the chance coincidences, by doing a "control" experiment under the same conditions as the actual experiment is more satisfactory than calculating the chance coincidences from the resolving time of the circuit.) In this way it was found that 20 of the 25 coincidences were genuine.

Bothe and Maier-Leibnitz also tried the effect of rotating one of the counters through 90° about the primary beam, when the number of coincidences fell to 9 ± 8. In another experiment they moved the scattering foil along the primary beam so as to bring θ, ϕ to values for which coincidences would not be expected on the particle theory, when the number of genuine coincidences again fell off.

A comparison of the number of observed coincidences with the number expected on the photon theory shows that the observed number is much too small, unless corrections due to the scattering of the recoil electrons inside the target, and inhomogeneity of the primary γ-rays is taken into account.

Similar experiments with a thorium source have been reported by Jacobsen[55], these also led to the conclusion that the coincidences were real.

In view of these results, there can now be little doubt of the correctness of a description of the Compton effect along the lines of the photon theory, as a discrete event involving particles, and the Bohr Kramers and Slater theory must be rejected. There is still however the difficulty of accounting for the negative result of Shankland's experiments.†

§ 5. ACKNOWLEDGMENT

In conclusion the author wishes to thank Prof. C. D. Ellis for the many helpful suggestions and criticisms he has made during the preparation of this report.

* It was shown by Williams[59] at this stage that Shankland's results were not in accord with the Uncertainty Principle.

† Since the above was written a great deal more evidence has accumulated in favour of this view. We may mention Shankland's own extension of his experiment[60], a repetition of the Compton-Simon experiment[61], and an ingenious experiment by Williams and Pickup[62] all of

REFERENCES

(1) Blackett and Lees. *Proc. roy. Soc.* A, **136**, 325 (1932), and references given there.
(2) Rutherford, Chadwick and Ellis. *Radiations from Radioactive Substances* (Cambridge, 1930).
(3) Wells. *Phys. Rev.* **47**, 573 (1935).
White. *Phys. Rev.* **47**, 591 (1935).
(4) Akiyama. *Jap. J. Phys.* **2**, 279 (1923).
(5) Joliot. *J. Phys.* **5**, 219 (1934).
(6) Lawrence, McMillan and Henderson. *Phys. Rev.* **47**, 273 (1935).
(7) Cockcroft and Lewis. *Proc. roy. Soc.* A, **154**, 246 (1936).
(8) Aston. *Nature*, Lond., **138**, 357, 613 (1936).
(9) Bainbridge and Jordan. *Phys. Rev.* **49**, 883 (1936).
(10) Bainbridge. *Phys. Rev.* **44**, 56 (1933).
(11) Henderson, Livingston and Lawrence. *Phys. Rev.* **46**, 38 (1934).
(12) Jordan and Bainbridge. *Phys. Rev.* **50**, 98 (1936).
(13) Oliphant, Kempton and Rutherford. *Proc. roy. Soc.* A, **149**, 406 (1935).
(14) Dempster. *Nature*, Lond., **138**, 120 (1936); **138**, 201 (1936).
(15) Dee and Walton. *Proc. roy. Soc.* A, **141**, 733 (1933).
(16) Dee. *Proc. roy. Soc.* A, **148**, 623 (1935).
(17) Taylor and Goldhaber. *Nature*, Lond., **135**, 341 (1935).
(18) Kurie. *Phys. Rev.* **47**, 97 (1935).
(19) Bonner and Brubaker. *Phys. Rev.* **49**, 223 (1936).
(20) For a summary of this work see Andrade, *Rep. Progr. Phys.* **1**, 269 (1934).
(21) *Report of the International Conference on Physics* (London, 1934).
(22) Ellis and Wooster. *Proc. roy. Soc.* A, **117**, 109 (1927).
(23) Meitner and Orthmann. *Z. Phys.* **60**, 143 (1930).
(24) Bohr. *J. Chem. Soc.* p. 349 (1932).
(25) Pauli, in discussions at Pasadena (1931), often quoted by other authors, first published in the *Reports of the Solvay Conference*, **7**, 324 (1934).
(26) Chadwick and Lea. *Proc. Camb. phil. Soc.* **30**, 59 (1934).
Nahmias. *Proc. Camb. phil. Soc.* **31**, 99 (1935).
(27) Fermi. *Z. Phys.* **88**, 161 (1934).
(28) Konopinski and Uhlenbeck. *Phys. Rev.* **48**, 7 (1935).
(29) Henderson. *Proc. roy. Soc.* A, **147**, 572 (1934).
(30) Sargent. *Proc. roy. Soc.* A, **139**, 659 (1933).
(31) Fowler, Delsasso and Lauritsen. *Phys. Rev.* **49**, 561 (1936).
(32) Kurie, Richardson and Paxton. *Phys. Rev.* **49**, 368 (1936).
(33) Crane, Delsasso, Fowler and Lauritsen. *Phys. Rev.* **47**, 971 (1935).
(34) Delsasso, Fowler and Lauritsen. *Phys. Rev.* **48**, 848 (1935).
(35) Bonner and Brubaker. *Phys. Rev.* **47**, 973 (1935).
(36) Tuve and Hafstad. *Phys. Rev.* **48**, 106 (1935).
(37) Newson. *Phys. Rev.* **48**, 790 (1935).
(38) Bonner, Delsasso, Fowler and Lauritsen. *Phys. Rev.* **49**, 203 (1936).
(39) Report of the B.A. Meeting. *Nature*, Lond., **136**, 468 (1935).
(40) Ising and Helde. *Nature*, Lond., **137**, 273 (1936).
(41) Alichanow, Alichanian and Dzelepow. *Z. Phys.* **93**, 350 (1935).
(42) Cockcroft, Gilbert and Walton. *Proc. roy. Soc.* A, **148**, 225 (1935).
(43) Haxel. *Z. Phys.* **93**, 400 (1935).
(44) Ellis and Henderson. *Proc. roy. Soc.* A, **152**, 714 (1935).
(45) Duncanson and Miller. *Proc. roy. Soc.* A, **146**, 396 (1934).
(46) Jaeckel. *Z. Phys.* **91**, 493 (1934).
(47) Leipunski. *Proc. Camb. phil. Soc.* **32**, 301 (1936).
(48) Bohr, Kramers and Slater. *Phil. Mag.* **47**, 785 (1924).
(49) von Baeyer. *Z. Phys.* **95**, 417 (1935).

(50) Bothe and Geiger. *Z. Phys.* **32**, 639 (1925).
(51) Compton and Simon. *Phys. Rev.* **26**, 289 (1925).
(52) Shankland. *Phys. Rev.* **49**, 8 (1936).
(53) Dirac. *Nature*, Lond., **137**, 298 (1936).
Peierls. *Nature*, Lond., **137**, 904 (1936).
(54) Bothe and Maier-Leibnitz. *Phys. Rev.* **50**, 187 (1936).
(55) Jacobsen. *Nature*, Lond., **138**, 25 (1936).
(56) Bonner and Brubaker. *Phys. Rev.* **50**, 781 (1936).
(57) Rumbaugh and Hafstad. *Phys. Rev.* **50**, 681 (1936).
(58) Bonner and Brubaker. *Phys. Rev.* **50**, 308 (1936).
(59) Williams. *Nature*, Lond., **137**, 614 (1936).
(60) Shankland. *Phys. Rev.* **50**, 571 (1936).
(61) Crane, Gaerttner and Turin. *Phys. Rev.* **50**, 302 (1936).
(62) Williams and Pickup. *Nature*, Lond., **138**, 461 (1936).

SOUND*

By E. G. RICHARDSON, B.A., Ph.D., D.Sc.

§ 1. FREE PROPAGATION: SUPERSONICS

The basic instruments for the study of the velocity and absorption of sound, viz. the Pierce interferometer, and the method suggested by Brillouin in which light is diffracted by the supersonic wave-pattern, realized simultaneously by Lucas and Biquard in France and Debye and Sears in America (cf. volume **1**, p. 68), continue to show advances in technique. The former type has become the established apparatus for gases while the latter is generally preferred for liquids, and has recently, as we shall see, been adapted to the study of supersonic velocities in solids. Yeagley[1] has built an elaborate supersonic generating and receiving system using three quartz-oscillators. The first is a frequency control and its oscillating circuit is resistance-capacity coupled to another containing the sending quartz. The object of this control is to hold the sending quartz at a steady frequency despite reactions due to gas pressure or other causes. The third quartz receives the radiation which has passed through the gas from the sender. Here the re-transformed electro-magnetic oscillations are amplified and measured on a peak voltmeter. The whole is contained in an enclosure maintained at a constant temperature. As in such a set-up it is impossible to screen the receiving circuit completely from electro-magnetic radiation springing from the sending circuit, a certain amount of direct coupling between the two circuits is permitted, producing a constant deflection B, in the absence of supersonic radiations, of the voltmeter. If the supersonic amplitude picked up by the receiving quartz is A, the deflection of the voltmeter under the combined action of the two radiations C, we have

$$C^2 = A^2 + B^2 + 2AB\cos\theta,$$

where θ is the phase-angle between the two radiations. A and θ will vary as the distance, x, between receiving and sending quartz is changed, but from a set of readings of C, both these quantities can be measured. Now as $A = A_0 e^{-\alpha x}$, the attenuation constant α can be obtained while a cycle of values of θ will correspond to a value of x equal to the wave-length; hence the velocity. It is of course essential for this method of working that multi-reflections between the quartz-oscillators should be ineffective on the receiver. This condition is secured by an ingenious tilting of the receiver with reference to the axis of the system, so that the second and third reflections interfere with each other on arrival at the plate. The author considers that the amplitude of higher-order multiple reflections is too small on arrival to affect the measurements. While still on the subject of the Pierce interferometer, we may note a convenient device due to Andrews[2] in which, though the

* The nomenclature of this Report is intended to conform to the *Glossary of Acoustical Terms and Definitions*, recently published by the British Standards Institution.

apparatus is of simpler form—single quartz and moving reflector—the mirror galvanometer which indicates the reactions on the quartz of the standing waves records on a moving film, whose motion is itself coupled to the reflector screw, thus giving a direct picture of the variations of current with distance of the reflector in true proportion.

Investigations with instruments of this type in 1935 fell mainly under two heads: (*a*) effects of pressure or temperature variation on supersonic dispersion in true gases; (*b*) the effects of admixture of foreign gases on dispersion. Last year initial measurements by Richards and his collaborators on the pressure and temperature effects were mentioned. Yeagley has used his apparatus to show that the attenuation in oxygen increases in rough proportion to the reciprocal of the pressure of the gas. Pumper [3] finds also that as the pressure is reduced the velocity rises, the amount being small in air but exceeding 2 per cent in carbon dioxide, over the range investigated. Both these workers used a single supersonic frequency. The work of Railston and the writer [4] covers a number of frequencies and pressures, but it has been found possible to plot the wave-lengths recorded for each gas on a single curve, if the quotient of frequency by pressure is used as parameter. In carbon dioxide and nitrous oxide the curves so obtained indicate a slight drop in velocity as the value of this parameter increases, followed by a steep rise and again a slight drop—recalling the familiar trend of velocity at a point of resonant dispersion in optics. The attenuation, measured by a hot-wire device, also increases as the reciprocal of the pressure, so that whatever the precise mechanism of the effect, location on the frequency scale seems to depend on the molecular volume of the gas. The effect of temperature, over a less extensive frequency range, has been examined by Penman [5]. His results indicate a dependence of the dispersion region on the absolute temperature directly, although he was not able to reduce all his measured wave-lengths in carbon dioxide to a single curve. Keeping to one frequency, however, he was able to trace out the familiar S-shaped curve—usually obtained at N.T.P. with *changing* frequency—by a gradual reduction of temperature. It should be added that neither these results nor those of Pearson (cf. § 3) are in agreement with the relaxation-time theory in its simple form, which demands a jump from the steady normal velocity (zero frequency) to a steady abnormal velocity (infinite frequency), when the time period of the radiation equals the time of relaxation of the gas.

The exponents of the relaxation-time theory are now engaged in testing mixtures of gases. We noted last year that Knudsen had observed an abnormal absorption at sonic frequencies when certain proportions of water vapour were mixed with air. The part which mixtures of such impurities play on the relaxation time of a gas has been elaborated by him and Kneser [6, 7] and by Marie Helme Wallmann [8]. It appears that the peak of absorption is shifted to higher frequencies as the concentration of the impurity is raised, indicating a reduction of the relaxation time. In oxygen, carbon dioxide and a number of other gases, the concentration at which the absorption is a maximum is proportional to the frequency, but the absolute value of the maximum is unaffected. These results indicate—to the physical chemist—that the reaction constant of the gas is strongly influenced by the impurities, on the

assumption that the absorption results from energy transfers between the original gas molecules. It is assumed that at any instant a gas is made up of a certain number of molecules which are active, i.e. able to vibrate when the radiation passes over them, and other momentarily passive molecules, the transitions between the two states being governed by reaction constants. The ease with which the wave is propagated will then depend on the average number of active molecules relative to the total number present. It is this activity which the impurity reduces—in direct proportion to its concentration in all cases except when the impurity is water vapour. This has been proved for nitrous oxide by Eucken and Jaacks[9]. Ishii[10] has found also that the velocity increases in humid air with the water-vapour pressure, and has found some evidence of dispersion, which may be the counterpart of Knudsen's absorption observations.

A good deal of theoretical discussion has arisen round the question of the "forbidden lines" in the diffraction of light by supersonic waves. Readers will remember that Brillouin predicted that a single ray of monochromatic light traversing a medium corrugated by sound waves would show two diffraction maxima on either side, due to the pattern imposed on the medium by the sound. As in many similar optical

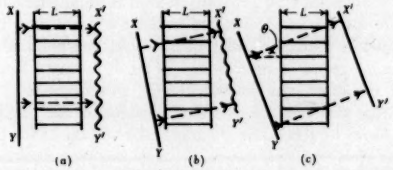


Figure 1. Oblique passage of light through liquid irradiated by supersonics (Raman and Nath).

effects, a large number of diffraction bands is in fact observed. By trying to produce mutual interference between these higher orders, Bär[11] has shown that they are due to multiple passages of the light from side to side of the trough containing the liquid medium. Raman and Nath[12] have extended the theory to give not merely the positions but also the intensities of the diffracted bands. They have also been able to explain one of the observations of Debye and Sears, i.e. that when the angle between the directions of light and supersonic waves is reduced from 90° downwards, the intensity of the diffracted light rays passes through a series of maxima and minima. For consider (figure 1 *a*) a plane wave-front of uniform illumination XY passing through a cell in which, at the instant pictured, the horizontal lines represent supersonic crests of compression, and therefore of supernormal refractive index to light. On emergence the wave-front $X'Y'$ will show maxima and minima of intensity as indicated. When the wave-front is oblique (*b*) so that the rays (shown by arrows) are inclined to the supersonic crests and troughs, the rays for the greater part of their paths through the cell lie in regions of almost normal refractive index. On emergence, then, the amplitude difference will be less marked, while when the inclination θ has reached that value shown in (*c*) the difference will have vanished. This will occur in fact whenever $\tan\theta = \lambda/L$, λ being the supersonic wave-length and L the width of the cell. The maxima between these positions of zero diffraction

actually decrease in magnitude as θ is increased. Another subsidiary effect has been pointed out by Ali[13], viz. changes in frequency of the emergent light from an original ν_0 to a value ν where $\nu = \nu_0 \pm mf$, f being the supersonic frequency and m the order of the diffracted maximum (*vide supra*). This has the physical consequence of resolving the spectral line into a doublet.

A number of applications (mainly of interest to the chemists) of the Brillouin method have been made to liquids, including the inevitable one on heavy water, by Bär[14], giving a slightly higher ratio of specific heats than for the common water. An interesting extension of this method to measure velocities of sound in solids has however been reported. Bez-Bardili[15] employs the opaque screen, pierced by a number of holes as source, in the same fashion as Bär and Meyer, and gets the spots drawn out into diffraction lines after passage of the light through the cell which contains xylol (cf. volume 1, figure 5). If a transparent solid in the form of a prism is placed in the cell, the refraction of the sound waves through it may be observed by the relative lengthening of the spots at different points in the field, and hence the velocity of sound in the solid calculated. In an alternative device (figure 2) a plate *PP* of the material, which this time need not be transparent, is placed in the

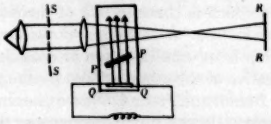


Figure 2. Total internal reflection of supersonics in a solid plate (Bez-Bardili).

cell interrupting the supersonic radiation from the quartz *QQ*. The diffraction of the light spots in the disc *SS* falling on the screen *RR* show that some supersonic radiation is reflected and some transmitted, but on turning the plate about a vertical axis the critical angle is reached at which total internal refraction of the supersonics ensues, and the unextended images of the spots beyond the plate demonstrate the disappearance of the transmitted wave. This angle is, of course, the critical angle (for supersonics) from xylol to the solid. The velocity in the liquid having already been found, that in the solid is obtained. The apparatus, therefore, works on the same principle as the familiar Wollaston apparatus for light waves. Sokoloff[16] allows the radiations from a quartz, under an applied potential of 2 or 3 KV, to pass through a solid, e.g. a piece of iron, and then into the liquid cell. The radiation which penetrates the iron is still intense enough to excite several orders of the Brillouin spectrum, and in this way he proposes to examine the solid for internal faults, bubbles, cracks, etc., which interrupt the transmitted radiation over part of the field, thus causing distortion of the spectrum. Hiedemann and Hoesch[17] have placed specimens of glass in contact with a piezo-electric source in a beam of polarized light between crossed Nicols (a method originally due to Schaefer and Bergmann), in this way picking out the nodes and antinodes by the light and shades in the resulting transmitted light. By separating out the effects due to longitudinal and

transverse vibrations, they have been able to calculate the elastic moduli of the glass and to observe the effect of mechanical compression on the material.

Before proceeding to discuss general supersonic effects, a special shape of piezo-electric oscillator, which should be particularly useful in producing many of these effects, will be described. Gruetzmacher[18] has cut a piece of quartz in the shape of a portion of a ring—the section of the annulus being a square—and placed it between similarly shaped electrodes. The quartz when oscillating sends radiation out radially, that on the outside of the ring diverging, while that on the inside converges on the centre, where a very intense local supersonic field is produced. In fact when the oscillator was set working in oil, a fountain 20 cm. high shot up from the centre of curvature!

Beginning with the heating effect, we note that Oyama[19] has carried out some useful work on the correlation between it and the physical properties of the fluid—in this case, aqueous salt solutions. He considers that the temperature rise is not determined by viscosity mainly, but is indeed proportional to the observed decrement of amplitude of the wave in its passage. Rather is the dissipation of energy characteristic of each liquid and but slightly dependent on circumstances such as temperature which may change the viscosity. Further, there is a range of frequency for maximum dissipation which is characteristic of the liquid. For pure water this range is from 700 to 800 kc./sec. A characteristic behaviour of a two-phase fluid under supersonic radiation is its emulsification or coagulation—one of which may be considered as the negative of the other, at least by the physicist, even if it is not so simple to the chemist. Brandt and Freund[20] have examined the rate of aggregation of aerosols—tobacco smokes—in the supersonic region. Here again there seems a most favourable frequency range for the coagulation of a given aerosol; 6·5 to 25 kc./sec. in the case of these smokes. Micro-photographs of the particles and measurements of the rate of sedimentation under gravity both showed a progressive increase in size as the radiation persisted. The rate of coagulation naturally increases with the amplitude of the source, and decreases as the mean free path between collisions increases. A more or less steady state is eventually reached, at which, owing to diminution of the numbers of particles in unit volume, the source, if maintained at constant amplitude, loses its effectiveness in favouring aggregation. The converse effect—disruption of large masses and dispersion in small particles—was one of the most striking features of the first supersonic effects shown by Wood and Loomis, and essentially requires very large amplitude. Bondy and Söllner[21] ascribe this phenomenon to the cavitation which results, particularly at the interface between the two phases, viz. mercury or oil and water. The collapse of water vapour into these cavities favours the dispersion of the more volatile phase into the less volatile. At the same time, it appears to the writer that there must be an aggregating influence at work, so that there should be a most favourable range of amplitude for either effect to predominate. This is a point which deserves investigation.

Claus[22] uses supersonics to enhance the dispersion of metals in suspension, by a novel process. The metal is produced by electrolysis in a voltameter, the kathode of which is directly coupled to a quartz-oscillator (200 to 2000 kc./sec.).

The vibration inhibits the usual deposition of the metal on the kathode; instead it is thrown back into the electrolyte in finely divided form. The degree of dispersion (average particle size) may be controlled partly through the frequency and amplitude of the supersonics and partly through alterations in the conditions of electrolysis (plate current density, etc.). He has also been able to disperse halogen silver in finely divided form in a photographic emulsion, whose sensitivity is said to be superior to that of the best commercial photographic plates having the same average grain size.

The efficacy of supersonics in peptizing the halogen silver in gelatine has been examined in further detail by his pupil Dangers[23]. The latter found that a previous temperature treatment of the emulsion was very important for ensuring the sensitivity of the emulsion and that the irradiation by supersonics was in the nature of a ripening process, taking the place of the chemical after-treatment usually employed.

In trying to produce mercury emulsions in water, Reggiani[24] notes that the presence of small quantities of electrolytes, whether acids or salts in the water, prevents high dispersion, but that the addition of protein nullifies the effect of the acid and restores the high degree of dispersion. This is apparently an effect of surface tension. This author with Marinesco[25] has demonstrated the formation of stationary waves in oil between the quartz and a photographic plate set at an angle of 45°, by the striations which are seen in the plate when developed. This points to a new supersonic effect, viz. a photochemical one. Finally the latter author[26] reports yet another use for supersonics, though perhaps one which might readily be expected. Detonating powders are set off under the action of the radiation. It is not clear what use can be made of this fact unless under certain circumstances such a method of deflagration were preferable to a blow or to the use of heat.

§ 2. VIBRATING SYSTEMS: SOLIOS

Pederson[27] has investigated the mathematics of a system under forcing in which the restoring force increases with the amplitude, i.e. one for which, with the usual convention of symbols, the equation is

$$m\ddot{y}+\mu\dot{y}+ky=F(1+Cy)\sin\omega t,$$

where C is a constant. Such a case occurs in practice when a loud-speaker has a moving coil working in a non-uniform magnetic field. The solution of the equation shows "resonance" at sub-harmonics of the forcing frequency; in particular the sub-octave is prominent. He has constructed a mechanical system, figure 3*a*, which illustrates the production of sub-harmonics very effectively. The idle-shaft H has two pulleys, one circular, and the other of such a shape that the distance of a point on its circumference from the axle increases from one side to the other. A cord round the former holds the mass M in suspension from the spring K. A cord round the latter is connected through the spring S to the crosshead D which is set in oscillation with the forcing frequency $\omega/2\pi$ by the rotation of the shaft B at ω rev./sec. Damping is introduced through the eddy currents in the aluminium

disc A moving between the poles of an electro-magnet. It will readily be appreciated that, as H oscillates, the force in the spring S is, owing to the peculiar shape of the eccentric pulley, of the type given by the right-hand side of the above equation, and this force is communicated to the mass M by the circular pulley. There is, actually, another condition for the production of sub-harmonics in a system of this type, as the mathematics shows. Transforming the equation to

$$\ddot{y} + 2\gamma\dot{y} + \omega_0^2 y = F_1 (1 + Cy) \sin \omega t,$$

we must have both $\omega = 2\omega_0$ and $|CF_1| = 4\gamma\omega_0 = 2\mu\omega_0/m$ for sub-harmonic resonance. In the experiment the damping (measured by log. dec. determination) was varied until resonance was obtained, when curves of the type shown in figure 3 *b* (continuous line) for the mass M were obtained, while D oscillated in simple harmonic motion (dotted line). Unfortunately, the actual damping was not independent of amplitude, so that a good check with theory was not obtained. (It is of interest to note, although the author is apparently unaware of the fact, that the sub-synchronous maintenance of forced vibrations was discovered by Raman and Dey in 1919.)

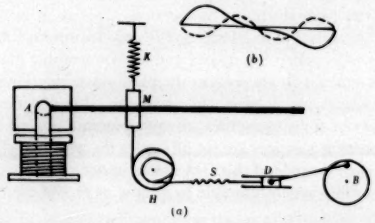


Figure 3. Apparatus to show sub-harmonic resonance (Pederson).

Ruedy[28] has done experiments on the forced longitudinal vibrations of nickel rods whose thickness is comparable with the wave-length of some of the more readily excited overtones. When the radial component is at or near resonance it absorbs some of the energy supplied and at the same time affects the velocity of the longitudinal vibrations, which has a related consequence, viz. the departure of the frequency of the nearby longitudinal overtones from their proper whole-number relation to the fundamental. An interesting case of a coupled system thus arises; in fact, the longitudinal velocity passes through the changes familiar in anomalous optical dispersion, as the rod passes through a radial resonance.

In connexion with the measurements of the amplitude of vibrating solids, two adaptations of the Bragg chattering method have been made. In the original method, the end of a springy strip was pushed up towards the vibrating body to such an extent that the spring was held off the vibrator permanently. Partridge[29] replaces the spring by a rigid pendulum of mass m, restoring force per unit displacement k, negligible damping, exerting a pressure p on a vibrating diaphragm. If the pivot

of the pendulum is moved a distance d behind the surface of the diaphragm, the equation of motion of the mass is

$$m\ddot{x}+k\,(d+x)=p.$$

Or substituting for x the motion of the diaphragm ($a \sin \omega t$) itself, we obtain

$$-m\omega^2 a \sin \omega t+k\,(d+a \sin \omega t)=p.$$

When chatter ceases, contact is no longer established ($p=0$) and oscillation of the pendulum stops ($\sin \omega t=1$). Then the required amplitude is given by

$$a=kd\,(m\omega^2-k)^{-1}=d\omega_0^2\,(\omega^2-\omega_0^2)^{-1},$$

where ω_0 is the natural pulsatance of the pendulum. As the latter quantity is generally small compared to ω, we can take $(\omega_0/\omega)^2\,d$ for the result, the quantity $(\omega/\omega_0)^2$ being the amplification factor. By making ω_0 small enough, amplitudes of the order 10^{-6} mm. may be measured. The method of Andrejev and Ugolnikov[30] may be called a subjective chattering method inasmuch as the finger is used to feel the amplitude. Considerable practice is of course necessary, in combination with independent objective measurements, before the amplitude can be thus gauged with any surety, and there is a considerable frequency dependence of the tactile sense. It is however a useful sense, though not possessed by all, for the qualitative estimation of amplitude, and the writer has had the experience—probably shared by others—of being able to feel the small amplitude of vibration communicated to objects in the house quite remotely and only mechanically connected to the domestic a.-c. supply, such as wooden lamp stands, mantelpiece ornaments, etc.

A number of works on the vibrations of strings are of mathematical interest only, but Ghosh[31] has drawn attention to the importance of the soundboard of the pianoforte in transforming the tone-quality of the struck string into aerial vibrations and Abbott[32] has initiated an elaborate investigation of the acoustical properties of violins. In order to secure identical conditions for all the violins tested—a precaution often neglected in such work—the instrument and recording microphones were placed at fixed points in a sound-lagged room (to obviate standing waves) and the bow-hairs in the form of an endless belt passed over the string at a fixed speed and pressure. The acoustical output of each of the open strings was compared, the mean deviation of power being taken as an index of the lack of tonal balance. The tonal quality was assessed by the relative strength of the lower partials, though the subjective meaning of such a test is not made clear. The impecunious musician will however be cheered to hear that a 25 dollar and a 10,000 dollar violin exhibited little difference in respect of either test!

Now that tuning-forks electro-magnetically maintained by the current from a microphone attached to the stem are becoming common in laboratories, it is useful to note that Lamson[33] has found such a fork made of steel, having a low temperature coefficient of frequency, to be a reliable standard source, its frequency and damping varying very little with any reasonable amplitude.

Some serviceable work has been done by A. B. Wood[34] and F. D. Smith[35] on the natural frequencies of circular plates and sheets of various materials. The method of excitation was in most cases by direct contact with nickel rods of appropriate

frequency, set in magneto-strictive oscillation, while to study the mode of vibration the Chladni method of sprinkled sand was applied. The small circular plates were tested at the natural frequency corresponding to one nodal circle. According to Poisson the radius r of this circle should be related to that of a thin plate, free at the edge, by the simple formula $r = ka$, where k is a constant of value 0·678. Wood does in fact find that this formula holds over quite a range of thickness and diameter. Only when the ratio of thickness to diameter reached one-fifth was there appreciable departure in the direction of decreasing k; at this stage the frequency also began to fall below that proper to a thin plate. This same method of excitation by contact with a magneto-strictive oscillator and making a Chladni plate of the substance, has led to a determination of the velocity of transverse waves in a number of commercial materials both alloys and insulators, to which the formula tested in the first research can be applied. The method is at its best in non-resonant materials such as cardboard, to which other ways cannot readily be devoted.

There are two papers of a musical interest which should here be catalogued. Giannini [36] continues his now extensive work on bells by directing attention to the importance of the position in a campanile at which the members of a carillon are hung. Contrary to the usual practice, the smaller bells should be nearer the ground, as the directive characteristics of radiation from a source become more pronounced at higher frequencies, and the polarity, in a vertical plane, of the radiation curve for a small bell hung high above the ground is such that the sound misses a listener at ground level. The general tendency to build tall towers should be checked, and a return to the lower levels of the famous Low Country carillons should be made. Obata and Tesima [37] have made records both of the vibrations of the twin membranes of a cylindrical drum, and also of the aerial sound waves produced by such an instrument. The point of greatest physical interest in their results relates to the relation between the vibrations of the skins (which should be tuned to the same pitch) at each end of the cylindrical column of air inside the drum. The air column couples the membranes in vibration; closely in a shallow drum; less tightly as the axis of the cylinder is lengthened. The interesting phenomenon of the intermittent action of two equal coupled pendulums is found again here in the alternate rise and fall of the vibration of each membrane, the energy being bandied about between them so that first one and then the other is at its maximum amplitude. Lengthening the cylinder, besides reducing the coupling, pulls the fundamental of the skins down—possibly by increasing the damping—though it appears not to affect the overtones. No suggestion is made to account for this curious observation.

§ 3. VIBRATING SYSTEMS: FLUIDS

Bubbles and drops continued to supply subjects for research in 1935. F. D. Smith [38] is interested in the vibrations of bubbles of gas in water from the aspect of the origin of the energy generated when intense sounds pass through water containing dissolved gas. We have above cited cavitation by supersonics as a contributory cause of emulsification. It is well known that intense supersonic radiation

can destroy organic matter in liquids through which it passes. This is ascribed to liberation of bubbles at the surface of the immersed object, the resonant vibration of the bubble involving a tearing action on the growth which eventually disintegrates it. Gorbatschew and Severny[39] arranged two liquid drops on the ends of fibres and investigated the forces between them in a hydrodynamic field. Both in continuous and alternating motion there is attraction between the drops. In the acoustical field, the drops partake of the motion and the attraction ensues as a result of this oscillation (Bjerknes effect). In practice, there is a frequency—possibly a resonance for the twin systems—in the supersonic gamut for which the attraction is so violent that the drops collide and coalesce.

G. B. Brown[40] has reported a very extensive and painstaking research on the sensitive jet, enriched with some beautiful photographs. His present work is principally concerned with the vortices, the passage of which up the smoky column is the ultimate cause of the breakdown from the steady form, and *ipso facto* of the sensitivity to vibration. The observations were easier to interpret by virtue of the fact that the slit from which the jet emerged was linear, instead of circular, as in the earlier work. Cinematograph films of the motion enabled the progress of the vortices up the jet to be followed. It is interesting that although the author finds approximate agreement between the experimental values of the ratio velocity of vortices : velocity of jet, and that calculated by Rayleigh (i.e. 1/2) when the slit is wide, yet narrow jets exhibited much lower ratios. This discrepancy between practice and theory is commonly found in related phenomena (aeolian tones, edge tones, Kundt's tube, etc.) when the orifice or obstacle falls below $\frac{1}{2}$ mm. in width. A small increase of vortex velocity with frequency was also found, but an important feature of all the films was that the rate ω of rotation of the vortex filaments was directly proportional to the frequency f of the sensitizing sound and independent of jet size or configuration. The author therefore concludes that it is this specified rate of angular rotation which characterizes the sensitivity of the jet, i.e. that for which $\omega = \pi f$ radians per sec. He has also shown that the origin of the disturbance is at the orifice itself, since, if this part is shielded from the sound, the jet becomes insensitive.

Under the title of "Flammes Excitatrices", Carrière[41] has published a paper on some experiments with singing flames in which, although a great number of observations have been made, it is difficult to sort out any facts of outstanding importance or novelty. Seiberth[42] also reports a fact well known to organ builders, viz. that two organ pipes, which, sounded alone, have slightly different frequencies, will pull into unison when sounded together. The phenomenon appears, however, to have escaped investigation by scientists, until the present time.

Passing on now to Kundt's tube phenomena, we observe the Dvorak method of a thin layer of oil applied to the walls and set into thin films traversing the tube at the nodes gives good wave-length measurements in the supersonic region (Brandt and Freund[43]). Pearson[44], who has used smoke in the same frequency scope, finds marked dispersion of the velocity in air. He observes, in fact, a rise of velocity (reduced to 0°) to a maximum of 333 m./sec. at 310 kc./sec. followed by a fall to

normal at 500 kc./sec. and another smaller hump at 630 kc./sec. This is again a dispersion of the type which is not satisfied by the relaxation theory in its present form (*vide supra*, § 1). The two peaks are ascribed to the two principal constituents, nitrogen and oxygen respectively, of the air. It was observed that a ring of smoke adhered to the walls of the tube at the nodes after irradiation for a short time, and photographs of these are reproduced from which the wave-lengths were measured. A water jacket completely encircling the tube was a necessary refinement, omitted by many experimenters with illuminated Kundt's tubes, which enabled convection currents to be precluded.

The well-known Kirchhoff formula for the velocity of sound in narrow tubes is not above suspicion, if one may judge from the number of research workers who still make experiments to check it. Using a tube of 6 mm. diameter with a thermophone as source and a pressure-gradient microphone in its neighbourhood as detector, Waetzmann and Keibs[45] measured the acoustic impedance of the tube at frequencies between 400 and 1400 c./sec., and hence calculated the damping coefficient. Satisfactory agreement with Kirchhoff is found. (These authors are actually concerned with the method of measuring the impedance of the ear-drum by the attachment of a tube of known acoustic characteristics, cf. volume 1, p. 84. It is of course necessary to know the acoustic behaviour of the tube before that of the ear can be deduced.) Norton[46] has pursued the same question into the supersonic region. Making use of the fact that thin nickel rods to fit narrow tubes may be set into magneto-strictive oscillation, this author constructs a Pierce interferometer in tubes $\frac{1}{2}$ in. to $\frac{3}{8}$ in. in diameter ($2r$) containing dry air freed from carbon dioxide. The frequency ranges from 5 to 110 kc./sec. and the velocity satisfies the formula

$$V = V_0 [1 - \gamma'/2r (\pi f)^{\frac{1}{2}}],$$

where V_0 is the free-space velocity. The value of γ' which Helmholtz deduced by considering the part played in dissipation by true viscosity and by heat conduction is actually less than the theoretical value (0·47 instead of 0·54), and the author makes a number of suggestions to account for this, of which the most plausible is that some of the heat produced by the passage of the waves may be conducted into and through the material of the tube, whereas Helmholtz pictured it as remaining in the gas.

Voelcker[47] on the other hand, investigating the damping in hollow spherical resonators filled with air or carbon dioxide, has obtained good agreement with theoretical values of the damping, provided oscillation of the walls is prevented.

It is a feature of the mathematics of the propagation of intense sound waves that some of the energy from a pure tone source should be translated into the second harmonic in the medium through which it is propagated, and Thuras, Jenkins and O'Neill[48] have successfully demonstrated and measured this intruding harmonic in the passage of sound through loud-speaker trumpets, both cylindrical and exponential. Goldstein and McLachlan[49] have also made a significant contribution to the theory of this problem, which becomes important now that public address systems dissipate as much as 500 watts of (electrical) power.

There are two instruments of great importance in the measurement of the amplitude of stationary waves, viz. the Dvorak disc, in which the pressure of the acoustic radiation on the disc is measured on a torsion balance, and the Rayleigh disc, in which the twist tending to set an obliquely suspended disc square-on to the wave-front is measured. In the simple theory of these devices, two by no means negligible factors were ignored, i.e. (1) the diffraction of the radiation by a small disc and its effect on the sound field, (2) the fact that the disc itself tends to swing with the field. The necessary corrections have been dealt with theoretically by King[50], but A. B. Wood[51] has independently examined both in theory and in experiment the magnitude of the second more considerable factor. The usual formula for the torque T on a disc of radius a, set at an angle θ to the direction of propagation of waves in a medium of density ρ, i.e.

$$T=[\tfrac{4}{3}\rho a^3]\,\bar{v}^2 \sin 2\theta \qquad \text{.....(1),}$$

must be multiplied by the inertia factor $(1-\beta)^2$. Here $\bar{v}$ is the velocity-amplitude in the fluid, $\bar{u}$ that of the disc vibrating in sympathy, $\beta=\bar{u}/\bar{v}$. Thus the relative velocity of fluid and disc is $\bar{v}\,(1-\beta)$, and this term should replace $\bar{v}$ in (1). The term in square brackets is the equivalent fluid load on a thin elliptical disc set at 45° to a stream. If M is the mass of the disc, m that of the fluid displaced by the disc (in the Archimedean sense), the ratio of velocities,

$$\beta=\frac{m+\frac{4}{3}\rho a^3}{M+\frac{4}{3}\rho a^3} \qquad \text{.....(2),}$$

also the intensity of the sound

$$I=\rho c\bar{v}^2=\frac{3}{4}\frac{cT}{a^3(1-\beta)^2} \qquad \text{.....(3),}$$

where c is the velocity of sound in the fluid. In this equation (due to Wood), every quantity can be measured experimentally (T in terms of the twist on the torsion head). In the attendant experiments, the author used a family of discs of the same radius but different mass—by varying the thickness and material—in water at a point in the sound field of unvarying intensity. Equation (3) then shows that T will be proportional to $(1-\beta)^2$; or the square root of the deflection,

$$\delta^{\frac{1}{2}}=\kappa'(1-\beta),$$

where κ is an apparatus constant. Writing (2) in the form

$$\tfrac{4}{3}\rho a^3=\kappa\,(M-m)\,\delta^{-\frac{1}{2}}-M \qquad \text{.....(4),}$$

Wood shows that this equation is satisfied by the experimental results, for a plot of $(M-m)\,\delta^{-\frac{1}{2}}$ against M gives a series of straight lines, whose intercept on the axis of M is constant and equal to $\frac{4}{3}\rho a^3/\kappa$. The formula (3) may then safely be used for measuring sound fields, and a set-up for this purpose is described. The possibility of flexural vibrations of the disc is considered in the light of the earlier paper[35], but it is shown that a metal disc 2 cm. in diameter and 2 mm. thick will be unaffected by resonance, at least in the audible region.

§ 4. ACOUSTIC IMPEDANCE AND PROPERTIES OF MATERIALS

During the year under review, both Hurst[52] and Erwin Meyer[53] have independently treated the problem of transmission through a manifold partition, and each has brought out the filter characteristics of this type of sound insulator. Hurst develops from first principles the theory of transmission through a series of plane infinite rigid partitions with constant separation, and shows that as the frequency of the oncoming sound is raised the system exhibits a series of alternate attenuation and transmission bands, the high pitch terminals of the latter having heads like band spectra corresponding to the condition $l=j\lambda/2$, where l is the spacing and j an integer. The latter author is more concerned with the frequency limits and the actual value of the attenuation which such a filter can be expected to exhibit in the lower part of the gamut, i.e. when the spacing l is small compared to the wave-length, the obverse condition to that postulated by the Canadian scientist. With this proviso, the system will form a high pass filter of limiting pulsatance $\omega=2c\,(lm/\rho)^{-\frac{1}{2}}$, m being the mass per unit area of the partition and ρ the density of the air. Experiments with 5, 10, 15 element model systems made of cellophane or cardboard confirmed this view, at least in its general features, small discrepancies in the expected attenuation being ascribed to resonant drumming of the air pockets. This could of course be reduced by straw or similar packing. To exhibit the efficacy of this class of insulator a full-size fourfold wall of 50 kg./m.^{2} and total depth 45 cm. was built and found to have a mean insulation of 55 db., as good as that of a single wall 50 cm. thick and 1000 kg./m.^{2}—a big saving in material.

In the domain of impedance measurement there is only one paper to note. Sivian[54] measures the reactance and resistance for circular orifices of diameters 1 cm. to 0·3 mm. in metal plates. The experimental arrangement is similar in principle to that of Robinson (cf. volume **2**, p. 178). The orifice divides a cylindrical chamber, backed by a loud-speaker, into two parts. By means of a microphone and search tube the pressure is measured in amplitude and phase just in front of the orifice and behind in the Helmholtz resonator formed by this portion of the chamber. The impedance of the orifice is supposed to be divided into two parts, viz.

$$Z=Z_1+Z_2=(R_1+iX_1)+(R_2+iX_2),$$

where the subscript 1 refers to the internal impedance of the orifice—computed from the Kirchhoff formula for tubes—and 2 to the associated disturbance on the sound field—the effect usually summed up as end-correction. The measurements (of which details are lacking) gave the total impedance, whence X_2 and R_2 were deduced. The reactance was found to be substantially independent of amplitude, and a formula for computing it is given, although as one of the assumptions made is the notoriously incorrect one of constant velocity-amplitude across the orifice, it is doubtful whether this formula will be helpful. The resistance term, however, increases with the amplitude, but has a limiting low-velocity value. The author suggests two causes for this. It may be due to turbulence, for which he reckons the lowest value of Reynolds' number quoted by authority is 1200. (We may remark

that Griffiths and Porter some years ago found the critical Reynolds' number to be 240 in one-way flow through such an orifice.) As, however, deviations from the low-velocity constant of resistance are found well below 300, the author thinks these may rather be due to losses of kinetic energy passing through the jet. The writer would suggest that it is just that non-uniformity of a.-c. flow through such an orifice cited above that is the cause of the increasing resistance with amplitude. Traverses with a hot-wire probe would show the author that the amplitude increases considerably in the outer regions of the aperture, dropping suddenly to zero at the walls. The magnitude of this annular effect increases with velocity-amplitude, thus enhancing the velocity gradient at the confines of the orifice, and *ipso facto* the viscous damping.

There have been many attempts—the first by Rayleigh—to draft a workable theory of the absorption by porous materials. The subject, of course, bristles with difficulties, and no progress can be made unless the material is idealized almost beyond recognition. The favourite conceit is to picture the substance as a series of narrow tubes having axes perpendicular to the exposed surface. Incidentally, the same assumption is made in calculating the rate of percolation of water through the soil, but the two problems are not analogous; alternating flow must not be confused with continuous flow. Cremer[55] elaborates the distinction between the resistance to continuous flow and that of alternating flow of air in such an ideal material. For calculating the resistance it is necessary to compute the mean velocity across the section. In the former case, this is the mean height of the parabola, but in the latter—again owing to skin effect—the profile is blunter and the mean level higher. The effect is markedly dependent on frequency. Setting $\delta p/\delta x = \mu\bar{v} + \rho\dot{v}$, where μ is a coefficient of resistance and ρ the density (inertia coefficient), we get with simple harmonic variation of v

$$-\delta p/\delta x = (\mu + i\rho\omega)\,\bar{v},$$

a formula familiar (in other symbols) to electrical engineers as the cable formula. Whereas μ and ρ are constants, if the parabolic profile is assumed, they are found to increase with the square-root of the frequency, when the true velocity-distribution is interposed. Another dubious point is whether one should assume the adiabatic or isothermal gas relation when dealing with these narrow tubes. The author does not advance arguments *pro* and *contra*, but merely indicates that the latter law would produce a $\gamma^{\frac{1}{2}} : 1$ increase in the apparent porosity of the material, in the sense of a correspondingly lower resistance. He points out that the rectitude of his theoretical assumptions cannot yet be tested against practice since the variation in measured resistance coefficients varies between 1 and 18,000 as between workers in different acoustical laboratories!

§ 5. SUBJECTIVE ASPECTS OF SOUND

It used to be thought that the cavity in the pharynx could, by its change of shape in the production of different vowels, react upon the vocal cords so that not merely the sounds of speech themselves, but the actual form of vibration of the vocal

cords, varied between one vowel and another. In experiments on excized throats of calves and men in which the vocal cords were tensioned and air blown between them, W. Trendelenburg and Wullstein[56] have disproved this. Apparently the vocal cords as a party to the coupled system are so "strong", relative to the air cavities of the mouth, nose and pharynx, that their vibrations cannot be appreciably altered by changes in the size or shape of these resonators. In this respect the voice behaves in an analogous manner to those coupled systems; string + belly, reed + pipe, in the violin and oboe, respectively.

The only other work of marked novelty on the voice is that by Wolf, Stanley and Sette[57] on some aspects of the singing voice, particularly in regard to the volume of breath used and the fluctuations of amplitude and frequency shown in records of the *vibrato*. They find that loudness is not necessarily accompanied by a high rate of breath expulsion, in fact, a well-trained singer uses little more breath in passing from *pianissimo* to *fortissimo*. As a matter of fact there is a slight fall in the rate between *pianissimo* and *mezzoforte* or in passing from low to medium pitch, with a corresponding slight rise to the loudest or highest notes. The *vibrato* is most pleasing to listen to when it involves fluctuations of about 6 periods per sec. Although some singers profess to be able to colour their voices in various ways so that we hear of dark, normal and white intonations—why not yellow or green?—actually this change of quality is performed at the expense of a substitute vowel. Thus a singer, requested to sing the word "pa" in these three intonations just instanced, actually produced in succession, "paw", "pah" and "pa(t)". Tolmie[58] has given a mathematical analysis of the *vibrato*.

Scales of loudness have come into the melting pot during 1935. Fletcher[59] in a lecture to the Franklin Institute, gave a good résumé of work on the inter-relation of pitch, loudness and timbre, carried out at the Bell Telephone Laboratories. That part of the lecture which dealt with their method of measuring loudness at the Bell Laboratories is of considerable interest. The experiments in essence involve a subjective judgment of "twice as loud", although it is not certain that this ratio possesses the same significance as the octave does in pitch perception. The first experiment deals with monaural versus binaural hearing. If the observer hears first with both ears and then with one only, it is assumed that the loudness has decreased to one half. The r.-m.-s. sound pressure at the single open ear being p_1, it is necessary to raise it to p_2 before the loudness seems equal to what it was with both ears open. With a number of listeners, corresponding pairs of values of p_1 and p_2 were measured, and are shown as o on figure 4*a*. In the second experiment a 1000-cycle reference tone is balanced at r.-m.-s. pressure p_1 against a pure tone of remote frequency until they sound equally loud. They are then sounded together, and again it is assumed that the loudness is doubled. The 1000-cycle tone sounding alone is now raised to pressure p_2 until it sounds equally loud to the combination. Such points are marked X on the figure. In the third experiment, which has been used by four different sets of observers, although only three are shown in the graph, the reference tone alone is listened to at pressure p_1, and then raised to p_2, such that the loudness appears doubled. The consistency in the results obtained by

these different methods is considered sufficient to warrant a loudness scale, based on successive doubling. Thus we can calculate from this figure the relative pressure increases corresponding to increases of loudness $= 2^n$. This has been done, taking a pressure ratio of 100 to correspond arbitrarily to an increase of loudness of 1000 times ($n = 3$), and the result shown by the curve on figure 4 (*b*). It will be noted that in accordance with the well-known relation between stimulus and sensation, the scales have been plotted logarithmically, and the straight lines obtained—at least over the range which matters in telephony, broadcasting, etc.—justify the choice of a logarithmic scale of loudness.

Such scales have also been reviewed by Abbott[60], who favours the retention of at least two of the many scales which have been proposed at different times. It will be noticed that the above scale—the Fletcher-Munson scale—is arbitrary in

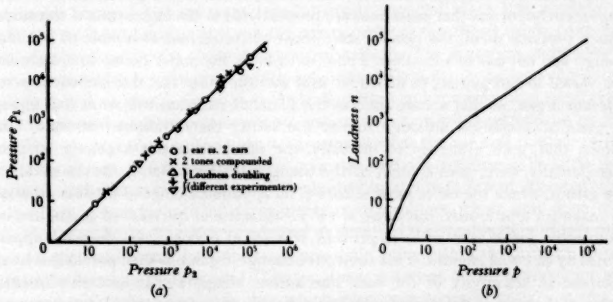


Figure 4. Experimental basis of a loudness scale (after Fletcher).

two respects, (1) the choice of the 1000-cycle reference tone, (2) the fixing of the reference threshold pressure at 2×10^{-4} bar, irrespective of frequency. Since, in fact, the ear is not equally sensitive at all audible frequencies, a higher level on the scale will not necessarily imply a greater loudness; thus "50 db. at 1000 cycles is louder than 60 db. [on this scale] at 100 cycles because the ear is more sensitive to the higher frequency." This state of affairs, though familiar to the psychologist, is confusing to the engineer, still more so to the cinema proprietor to whom the "acoustics dealer" wants to sell his sound absorbent. For such cases it would be better to use the minimum audibility at each pitch as the reference pressure, and keep the more coherent Fletcher-Munson scale for physicists and psychologists. It is important, however, as Dr A. H. Davis has emphasized, to *specify the threshold* in every case, if we are to understand each other's scales.

A discussion at one of the Physical Society's meetings during the session, initiated by a paper by Campbell and Marris[61], which the authors tried to keep within the province of physics, contrarily enough proved to be a field day for psychologists. The sponsors of the paper were not concerned with what scale

worked best in practice but questioned whether any scale at all could be built up on the facts as we now know them. In marshalling the facts of audition—although a number of these were discredited by various speakers—and examining their relation to the mental estimate of loudness, they have done considerable service in clearing the air of many false premises which mar the efforts of a number of workers in this field. The impression gained, however, was that they endeavoured to establish loudness-measurement on the same basis as a purely physical measurement, and failed in that "they seemed to regard the sense organs as physical instruments in which, when you apply a certain stimulus at one end, you will always get the same response at the other"—to use the words of one speaker—that is, they failed to make due allowance for the influence of earlier stimuli on the response of the ear to the latest one. Another point in the discussion which deserves emphasis is the difference between a sensation and the stimulus which produces it. While it appears to be correct that sensations are proportional to the logarithms of the stimuli which produce them, the decibel scale must be recognized as a scale of stimulus-energy and not one of sensation. These seem to be the major points in a discussion too varied to summarize, to which we may add the other fact that the ear is a very tolerant organ, so that a scale such as the Fletcher-Munson will work well enough in practical acoustics, although it may not satisfy the fastidious. Stevens[62] has shown that, with a change of intensity, the mind associates a change of pitch. (Incidentally, there goes another of the Campbell-Marris facts!) In the middle of the gamut, where the ear is most sensitive, no apparent change of pitch is produced by making a tone louder, but down at 150 c./sec. a rise of intensity of 50 db. involves a subjective *fall* of as much as 12 per cent, whereas at 12,000 c./sec. the pitch appears to rise by an equal amount. This subjective change of pitch is then parallelled by the decrease in sensitivity of the ear. The author suggests that excessive intensity acting at the ends of the basilar membrane changes the spatial stimulus pattern on it. Montgomery[63] stresses the effect of the method used on the results in such sensitivity measurements; particularly has the type of judgment demanded of the patient an important effect on the results. When the step in intensity or pitch is reduced until the subject reports "no change", the criterion is most difficult to determine. A large number of observations on different people is required to get a good statistical average ear. One feels that telephone engineers are at last becoming convinced of the wisdom of the artifices which psychologists have used for years.

Measurements of the time for which a pure tone must continue in order that the ear shall be able to estimate its pitch are of some importance for the sensitivity of the ear; in fact, the course of the curve obtained by plotting such times against frequency bears a remarkable similarity to the pitch:sensitivity curve for the ear. Burck, Kotowski and Lichte[64], who have recently completed a set of such measurements, date such investigations back to 1873 when Mach found for a single tone of 128 c./sec. a duration of 4 to 5 vibrations sufficient for pitch estimation. Some psychologists opine that 2 vibrations are sufficient, but this surprising view is not confirmed by this new work. The time required for the ear to apprize a suddenly presented tone is nevertheless remarkably short. The authors have also experimented

with a new form of interrupted tone. Instead of employing the usual type in which a tone of constant frequency is broken at regular intervals by "silences", they interrupt a continuing sound (1000 c./sec.) by the sudden intrusion of one of another pitch for a short interval. As the pitch of the interloper passes over the audible scale of frequencies the minimum time for the intrusion to be detected varies in the same fashion and at approximately the same absolute values as the time of excitation in their first experiments. This might be reckoned as the time lag in the response of the aural system, but the matter is a little more complicated than that, for the interruption or onset of a simple tone of short duration is, as the Fourier analysis shows, associated with a number of overtone transients, whose inception—together with starting noises peculiar to the instruments themselves—is responsible for the plop or click which is associated with each new tone when listening to loudly played or highly amplified music or speech. In the Helmholtz theory of harmony, dissonance between two or more simultaneously sounded tones was ascribed to more or less rapid beating between the components of the chord, while concords were characterized by the absence of noticeable beats. The sensation of roughness in the ear may also be given, apart from intentional discords, in the loud reproductions of speech or music, due to non-linear distortion and consequential unpleasant beats between the inharmonic overtones and combination tones introduced by the amplifier. It is therefore profitable to investigate what the ear will tolerate in the way of roughness, and if possible to set up a standard of roughness. The latter task has been undertaken by von Bekesy,[65] who has to a certain extent repeated with modern equipment Helmholtz' experiments on this subject, which were done with the aid of a violinist. Von Bekesy adopted as standard the combination of two pure tones of 3000 and 3050 c./sec., the roughness scale being that of the pressure amplitude of the beat tone in dyne/cm.^{2}. The roughness sensation of this combination was compared with others and with that which even pure tones in certain ranges of frequency evoke, but the results must be regarded as tentative as much was found to depend on the personal tolerance of the subject. Further the roughness of a combination varies considerably with the absolute loudness of the component tones, instead of, as was hoped, on the frequency and beat-amplitude solely.

Kock[66] discusses the same and other related problems in terms of the uncertainty principle. This, in so far as it affects listening problems, has been stated by Stewart in the form $\Delta f \Delta t \approx 1$. Here Δf represents a change of frequency occurring in a time Δt, and the relation indicates that for a given time Δt the pitch of a tone cannot be ascertained with an accuracy greater than the frequency interval Δf which satisfies the above relation. Then, for a given time interval, the pitch of a tone in the treble can be determined with greater accuracy than one in the bass, while a note rich in high harmonics can *ceteris paribus* be sensed more readily than a pure tone. On a keyboard instrument, in which the pitch perforce changes by steps, the rapid playing of a chromatic scale will give the sense of a continuous rise or fall of tone, corresponding to *portamento* playing on a violin or trombone provided (*a*) the average pitch is not too high and (*b*) harmonics are absent. (Cinema organists do not need to be told of this effect!) In the case of the *tremolo* or frequency

vibrato, the principle of uncertainty indicates that there should be a minimum *tremolo* speed for the *tremolo* to be apparent as such. The oscillation about the standard frequency should be less rapid in the bass than in the treble. Incidentally this condition is usually assured on the stringed instruments by the increasing difficulty of the manual dexterity required to play a *tremolo* on a large stringed instrument. Another faculty required of the player on a stringed instrument is that of being able to tune his strings, although it is true that the player on a pianoforte usually carries out this task by proxy. There is another difference between tuning on the pianoforte and on the violin. The pianoforte is tuned to the equi-tempered scale, whereas the violinist is said to use the unadulterated scale. How far this difference is a real one is debateable. Probably, constant playing with the pianoforte makes the violinist, even when playing alone or with other strings, unconsciously adopt the equi-tempered scale. Grutzmacher and Lottermoser[67] have been making an examination of piano-tuners to find how accurately they could tune to the equi-tempered scale. Objective measurements of the frequencies of the strings after tuning showed that, as one might expect, only an approximation to the equi-tempered scale was reached. The maximum error was about 10 cents, and the minimum was usually reached in the second octave above middle C. As the tuner tunes mostly by extinguishing beats, the results on "roughness" obtained by von Bekesy (*vide supra*) should have a distinct bearing on the estimation of this faculty, viz. the ability to set an instrument in tune.

Finally, we note an anatomical investigation on the ear of a cat, which should do much to establish the resonance theory of hearing among such doubters as are still unconvinced. In some earlier experiments Weaver and Bray had reported that sounds delivered into the ear of a decerebrate cat could be picked up and amplified as electrical currents *of corresponding wave-form* in the auditory nerve, a result which seems to favour the telephone theory of hearing. Hallowell Davis[68] has shown recently, however, that this effect is in the main due to an electrical impulse generated in the cochlea and carried by ordinary electro-magnetic radiation into the amplifier. If this secondary effect is shielded by suitable electrodes, frequencies above 3000 per sec. cannot be recovered from the auditory nerve, while in the tracts of the mid-brain, 1000 c./sec. seems to be the upper limit, whereas a cat can hear at least up to 20,000 c./sec. The nerve fibres in fact become fatigued after a few impulses and unless rested—as they would be in low-frequency stimulation—cannot take up a new potential until the elapse of a refractory period. The fibres respond to high-frequency a.-c. only in part, or in asynchronous fashion. He concludes that 1000 per sec. is the limiting frequency of impulses which can be transmitted through the fibres of the auditory nerve, just as it is for other mammalian nerve fibres of similar anatomical characteristics. Analysis is therefore the function of the cochlea. "Differentiation of pitch depends on activity of one or another zone of the basilar membrane and consequent differences in the distribution of activity among the fibres of the auditory nerve. Loudness does not depend on the number of impulses per second travelling up the auditory nerve, but simply on the number of active fibres, irrespective of the frequency of impulses in those fibres." These assertions

are supported by audiograms of cats which had defective hearing in some part of the audible range and in which post-mortem examination showed damage to a definite region of the basilar membrane.

§ 6. TECHNOLOGY

Minor improvements in recording and analysing apparatus are continually being published, but we have room here only to record those devices which involve major advances or specially commend themselves by their convenience in handling. Freystedt[69] constructs a frequency analyser using a number of electrical filters distributed over the audible range. At about 3 filters per octave, 22 suffice to cover the range 40 to 5000 c./sec. The amplified electric current from the microphone passes through all these in parallel, but a commutator picks up from each filter in turn, doing the round in the same time that a sweep circuit moves the electron spot in a cathode-ray oscillograph across the screen (about 1/10 sec.). Thus within the resolving power of each filter the spectrum of the sound is delineated on the screen and the whole analysis performed in this short time. Each filter consists of a pair of close-coupled resonant circuits, which gives sharper tuning than if a single circuit is used. At the frequencies where the filters overlap, the amplitude is 0·7 of the peak value, and at each peak, the amplitude of the neighbouring filters is 0·1 of the peak. Examples of spectra obtained by this apparatus are given by the author. Trendelenburg and Franz[70] and Vierling[71] also use an oscillograph and bank of filters (only one to each octave) but in place of the spectra of Freystedt they reproduce the wave-form of the filtered sound for each octave in turn. In this way they are able to study the rise and fall of each octave of overtones in the enunciation of short syllables and in the *Einschwingen*• of musical instruments. Erwin Meyer and Keidel[72] describe an ingenious decibel-meter which should find many uses in technology. As is well known, it is the custom of acoustical engineers to take the logarithm of the voltage generated in a microphone of uniform frequency characteristic as a measure of the loudness of the sound, although it is still contested with some justice that this is not a true measure of subjective loudness (cf. § 6). The poor engineer is naturally desirous of having an instrument to give him an objective reading of loudness, and so takes the logarithm of the response as a suitable compromise. Instruments employing variable-mu valves have already been designed for this purpose, but in the opinion of the authors fall short of their objectives. They accordingly tackle the problem in a more fundamental way and use a potentiometer in which the voltage tapped off bears an exponential ratio to the movement of the jockey. This is done by means of a liquid rheostat (with conductivity-water as electrolyte in a shallow waxen trough) to either end XY of which the a.-c. from the microphone is led, figure 5. The dipper D, carried on an extension of the pointer of the d.-c. ammeter feeds the proportion DX/XY of this voltage V to an amplifier from which the amplified current, after rectification, is passed to the same ammeter, A. Initially, the pointer of the ammeter is at zero and the dipper at the vertex X of the

• We do not appear to have a neat word for this in English. "Starting noise" is expressive but uncouth.

exponential flare. As soon as the microphone picks up sound, the pointer moves over the dial and the dipper moves over the wide end of the vessel on an arc of a circle keeping near its inner edge, which has been made to conform to this curve by twisting the axis of the vessel. As soon as this movement occurs the supply to the amplifier is reduced, and equilibrium will be reached when the dipper comes to rest at a point in the liquid such that its movement DX entails a certain steady rectified current I. For I to be a decibel measurement we require the microphone voltage V to be proportional to exp I, i.e. to exp $[(DY/XY)\,V]$. This is assured in this type of rheostat, at least approximately, since the conductance of any portion of it is proportional to the logarithm of the axial intercept. The authors have calibrated their instrument by applying small alternating voltages to XY and find that the relation $I \propto \log V$ holds between 2 millivolts and 20 volts, that is for a range of ten-thousandfold or (in terms of 2 millivolts as threshold potential) 80 db., which is quite sufficient for ordinary sound measurements. Examples are given of the application of the instrument to (*a*) frequency-response curves of microphones and loud-speakers, (*b*) sound transmission of walls over a range of

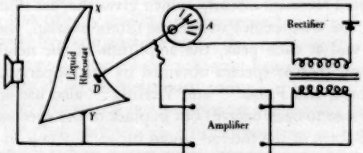


Figure 5. Logarithmic loudness-recorder (Meyer and Keidel).

frequencies, (*c*) reverberation of rooms, (*d*) decay curves of tuning forks, all of which can of course be quickly read off in decibels on the instrument. Wente, Bedell and Swartzel[73] describe another instrument for making these types of record with great dexterity. Their high-speed level-recorder is of the category in which the potentiometer slider is moved by logarithmic *steps* under the action of a magnetic clutch. The clutches are operated by the input voltage, the higher this voltage is, the further the slider is moved. Similar types of record to those noticed above are given by these authors and also by Wolf and Sette[74] using the same apparatus, which, though better adapted to quick operation, does not seem to give as much detail as the slower apparatus of Meyer and Keidel, besides being, one would imagine, more complicated and expensive.

For some purposes, e.g. when the intensity of a noise is rapidly changing, it is desirable to integrate the total noise picked up by a microphone over a long period. For this purpose, Haynes[75] replaces the final current reading instrument on such a recorder by a watt-hour meter of the type which measures the energy used in terms of the total rotation of a disc, damped by eddy currents.

Perhaps the simplest method of recording reverberation times, though not the most accurate, is to clap the hands together and listen to the dying away of the

resulting sound. Cotton[76] has improved on this, while retaining the same principle. A neon lamp is whirled round a circle and lights up whenever the gain of the microphone and amplifier rises above a pre-arranged value determined approximately by minimum audibility. The source is a loud-speaker which is actuated once in every revolution of the disc carrying the neon lamp, when contacts on the disc cause a condenser to be discharged through the loud-speaker, imitating a hand-clap. The length of the illuminated arc, which must not, of course, exceed the whole circumference, indicates the reverberation time. The disc may be revolved at one of three standard rates, 120, 60 or 15 rev./min., depending on the reverberant characteristics of the room under investigation.

Although great advances have been made in the technique of the measurement of reverberation it is still doubtful whether measurements in a reverberation chamber will give us an absolute measure of the absorption per unit area of materials introduced therein, which is a pity, since such a method most closely touches the practical requirements of building acoustics. We noted last year that considerable discrepancies were noteworthy between values for the coefficient of the material Celotex, when this material was used in the various proprietary and state acoustic laboratories. P. Sabine[77] then ascribed these discrepancies to differences in the area of the sheet used, but a re-examination of the data now suggests to his mind that it is the area relative to the total surface of the laboratory and not the absolute area of the specimen that is the significant factor. Although these deviations may be annoying to a physicist, to the occupant of the premises to which absorbent material is applied they do not amount to much, the ear being very tolerant of considerable variations in this respect.

Three Japanese scientists[78] have drawn attention to an intriguing case of constructional echo in a temple in Nikko, known as the "singing dragon". The ceiling is concave and of lacquered wood and if one takes up a position under the head of a dragon painted on the ceiling and claps the hands, "he hears a series of rattling or trembling sounds with successively decaying intensities lasting about half a second, as if the dragon answered the call of the clapping hands with a peculiar rattling murmur, causing sometimes uneasy sensations on the part of the worshipper". The phenomenon seems to be conditioned by successive reflections between the ceiling and the floor which is also of polished wood, and as the position determined by the dragon's head is near the centre of curvature of the ceiling reflections take place as it were between a concave mirror and a plane mirror at the centre of curvature of the former. Oscillograph records, not in the temple itself, but under a concrete arch, which exhibited the same echo, confirmed this view. The rarity of the phenomenon is no doubt due to the fact that the centre of curvature of most barrel vaults or arches lies above the ground, though the writer has observed the effect under a concrete arch from a boat on the river beneath.

REFERENCES

(1) YEAGLEY. *Rev. sci. Instrum.* **6**, 148 (1935).
(2) ANDREWS. *Rev. sci. Instrum.* **6**, 167 (1935).
(3) PUMPER. *Phys. Z. Sowjet.* **8**, 300 (1935).
(4) RAILSTON and RICHARDSON. *Proc. phys. Soc.* **47**, 533 (1935).
(5) PENMAN. *Proc. phys. Soc.* **47**, 543 (1935).
(6) KNUDSEN. *J. acoust. Soc. Amer.* **6**, 199 (1935).
(7) KNESER and KNUDSEN. *Ann. Phys.*, Lpz., **21**, 682 (1935).
(8) WALLMANN. *Ann. Phys.*, Lpz., **21**, 671 (1935).
(9) EUCKEN and JAACKS. *Z. phys. Chem.* **30**, 85 (1935).
(10) ISHII. *Sci. Pap. Inst. phys. chem. Res.*, Tokyo, **26**, 201 (1935).
(11) BÄR. *Helv. phys. Acta*, **8**, 591 (1935).
(12) RAMAN and NATH. *Proc. Ind. Acad. Sci.* **2**, 406 (1935).
(13) ALI. *Helv. phys. Acta*, **8**, 502 (1935).
(14) BÄR. *Helv. phys. Acta*, **8**, 500 (1935).
(15) BEZ-BARDILI. *Z. Phys.* **96**, 761 (1935); *Phys. Z.* **36**, 20 (1935).
(16) SOKOLOFF. *Phys. Z.* **36**, 142 (1935).
(17) HIEDEMANN and HOESCH. *Z. Phys.* **96**, 268 (1935); **98**, 141 (1935).
(18) GRUETZMACHER. *Z. Phys.* **96**, 342 (1935).
(19) OYAMA. *J. Inst. elect. Engrs*, Japan, **55**, 985 (1935).
(20) BRANDT and FREUND. *Z. Phys.* **94**, 348 (1935).
(21) BONDY and SÖLLNER. *Trans. Faraday Soc.* **31**, 835 and 843 (1935).
(22) CLAUS. *Z. tech. Phys.* **16**, 80, 109 and 202 (1935).
(23) DANGERS. *Z. Phys.* **97**, 34 (1935).
(24) REGGIANI. *C.R. Acad. Sci.*, Paris, **200**, 123 (1935).
(25) REGGIANI and MARINESCO. *C.R. Acad. Sci.*, Paris, **200**, 548 (1935).
(26) MARINESCO. *C.R. Acad. Sci.*, Paris, **201**, 1187 (1935).
(27) PEDERSON. *J. acoust. Soc. Amer.* **6**, 227 (1935); **7**, 64 (1935).
(28) RUEDY. *Canad. J. Res.* **12**, 10 (1935).
(29) PARTRIDGE. *Phil. Mag.* **20**, 953 (1935).
(30) ANDREJEV and UGOLNIKOV. *C.R. Acad. Sci.*, U.S.S.R., **1**, 28 (1935).
(31) GHOSH. *J. acoust. Soc. Amer.* **7**, 21 (1935).
(32) ABBOTT. *J. acoust. Soc. Amer.* **7**, 111 (1935).
(33) LAMSON. *J. acoust. Soc. Amer.* **7**, 51 (1935).
(34) WOOD. *Proc. phys. Soc.* **47**, 794 (1935).
(35) WOOD and SMITH. *Proc. phys. Soc.* **47**, 149 (1935).
(36) GIANNINI. *J. acoust. Soc. Amer.* **7**, 135 and 293 (1935).
(37) OBATA and TESIMA. *J. acoust. Soc. Amer.* **6**, 267 (1935).
(38) SMITH. *Phil. Mag.* **19**, 1147 (1935).
(39) GORBATSCHEW and SEVERNY. *Kolloidzschr.* **73**, 146 (1935).
(40) BROWN. *Proc. phys. Soc.* **47**, 703 (1935).
(41) CARRIÈRE. *Rev. Acoust.* **4**, 149 (1935).
(42) SEIBERTH. *Hochfrequenztech. u. Elektroakust.* **45**, 148 (1935).
(43) BRANDT and FREUND. *Z. Phys.* **91**, 415 (1935).
(44) PEARSON. *Proc. phys. Soc.* **47**, 136 (1935).
(45) WAETZMANN and KEIBS. *Ann. Phys.*, Lpz., **22**, 247 (1935).
(46) NORTON. *J. acoust. Soc. Amer.* **7**, 16 (1935).
(47) VOELCKER. *Ann. Phys.*, Lpz., **24**, 361 (1935).
(48) THURAS, JENKINS and O'NEILL. *J. acoust. Soc. Amer.* **6**, 173 (1935).
(49) GOLDSTEIN and MCLACHLAN. *J. acoust. Soc. Amer.* **6**, 275 (1935).
(50) KING. *Proc. roy. Soc.* **153**, 1 and 17 (1935).
(51) WOOD. *Proc. phys. Soc.* **47**, 779 (1935).
(52) HURST. *Canad. J. Res.* **12**, 398 (1935).
(53) MEYER. *Elekt. Nachr.-Tech.* **12**, 393 (1935).

(54) SIVIAN. *J. acoust. Soc. Amer.* **7**, 94 (1935).
(55) CREMER. *Elekt. Nachr.-Tech.* **12**, 333 and 362 (1935).
(56) TRENDELENBURG and WULLSTEIN. *S.B. preuss. Akad. Wiss.* **21**, 399 (1935).
(57) WOLF, STANLEY and SETTE. *J. acoust. Soc. Amer.* **6**, 255 (1935).
(58) TOLMIE. *J. acoust. Soc. Amer.* **7**, 29 (1935).
(59) FLETCHER. *J. Franklin Inst.* **220**, 405 (1935).
(60) ABBOTT. *J. acoust. Soc. Amer.* **6**, 137 (1935).
(61) CAMPBELL and MARRIS. *Proc. phys. Soc.* **47**, 153 (1935).
(62) STEVENS. *J. acoust. Soc. Amer.* **6**, 150 (1935).
(63) MONTGOMERY. *J. acoust. Soc. Amer.* **7**, 39 (1935).
(64) BURKE, KOTOWSKI and LICHTE. *Elekt. Nachr.-Tech.* **12**, 278, 326 and 335 (1935).
(65) VON BEKESY. *Z. tech. Phys.* **16**, 276 (1935).
(66) KOCK. *J. acoust. Soc. Amer.* **6**, 56 (1935).
(67) GRUTZMACHER and LOTTERMOSER. *Phys. Z.* **36**, 903 (1935).
(68) DAVIS. *J. acoust. Soc. Amer.* **6**, 205 (1935).
(69) FREYSTEDT. *Z. tech. Phys.* **16**, 528 (1935).
(70) TRENDELENBURG and FRANZ. *Z. tech. Phys.* **16**, 513 (1935).
(71) VIERLING. *Z. tech. Phys.* **16**, 528 (1935).
(72) MEYER and KEIDEL. *Elekt. Nachr.-Tech.* **12**, 37 (1935).
(73) WENTE, BEDELL and SWARTZEL. *J. acoust. Soc. Amer.* **6**, 121 (1935).
(74) WOLF and SETTE. *J. acoust. Soc. Amer.* **6**, 160 (1935).
(75) HAYNES. *Rev. sci. Instrum.* **6**, 321 (1935).
(76) COTTON. *Rev. sci. Instrum.* **6**, 344 (1935).
(77) SABINE. *J. acoust. Soc. Amer.* **6**, 239 (1935).
(78) NAKAMARA, OBATA and KURIKARA. *Proc. imp. Acad. Japan*, **11**, 313 (1935).

THE MEASUREMENT OF NOISE

By G. W. C. KAYE, O.B.E., M.A., D.Sc.,
Superintendent of Physics Department, National Physical Laboratory,
Teddington, Middlesex

§ 1. INTRODUCTION

A GENERATION ago the physical study of sound was regarded as something of no great practical account. Its treatment at schools was perfunctory, while at the Universities it was compressed into the least possible number of lectures, fobbed off with stereotyped experiments based on the tuning fork, the Helmholtz resonator, the sensitive flame, the bird-call, the resonating air column and the Chladni plate.

The foundations were there, largely provided by the profound mathematical and experimental skill of the late Lord Rayleigh, but the trouble with acoustics was that it lacked accurate measuring devices and the stimulus which would have come had it been of notable service to the public.

The touchstone came from Fleming whose thermionic valve presently transformed nebulous acoustical appraisement into precise electrical measurement. The sensitive flame gave place to the microphone and the Helmholtz resonator was supplanted by the resonating electrical circuit. Broadcasting and the talking pictures followed, the whole-hearted interest of the public was permanently captured and, as was inevitable, there came to every branch of acoustical science major and long-awaited developments in technique and equipment which have in their turn facilitated and accelerated further progress. The terminology of acoustics, which had developed out of recognition, has received the attention of a Committee of the British Standards Institution, whose *Glossary of Acoustical Terms and Definitions* has recently been published. The glossary includes both well established and new terms in sound and vibration, wave propagation, hearing, noise, architectural acoustics, transmission systems, sound reproducers, microphones, miscellaneous instruments and music. The glossary should serve as an authoritative reference for teachers and students, technicians and others.

The measurement of the loudness of sound and noise forms an important section of the glossary—particularly in view of the growing force of public opinion on questions of noise and its abatement. The people are beginning to realize that they have the right to protest against unnecessary transport noises whether on the road, on the water, on the rail or in the air. In their homes and particularly in the case of flats they demand protection against the noises of their neighbours. They are beginning to question the rights of public or private authorities to generate excessive noises whether from permanent machinery and plant, or from temporary nuisances such as pneumatic drills or riveting operations.

§ 2. SOUND AND NOISE

It used to be contended by writers of text-books on physics that noise was sound in which rhythm or musical character was conspicuously absent, but this has become unsafe ground in view of the later developments of music. A listener is, moreover, greatly influenced by environment, background or force of association though no doubt there are people who would regard any sound made by somebody else as a noise, if not a nuisance. It is common experience, however, that even if we exclude the hypersensitive and the totally immune there are many border-line sounds which would lead to very divergent opinions among a collection of reasonable people. The appraisement of sound as noise is in fact a personal matter of opinion. The British Standard Glossary defines noise as "sound undesired by the recipient", and a definition on some such lines is likely to be accepted internationally in the near future.

As is well known, sounds and noises are commonly made up of prescribable mixtures of pure tones of various frequencies and intensities. These are objective features which lend themselves to instrumental measurement and analysis. But from the point of view of the listener, there are in addition subjective features which are of prime significance, namely pitch, timbre and loudness, and these are not readily assessable.

§ 3. FREQUENCY

The extreme frequency range of the ear extends from, say, 20–20,000 c./sec., the upper limit declining substantially with advancing years. Present-day technical acoustics commonly restricts itself however to a narrower range from, say, 50–10,000 cycles. Incidentally the low-frequency components of a sound usually impart volume and largely control the pitch, while the high-frequency components mostly provide the identifiable characteristics of the sound.

§ 4. INTENSITY

The intensity range of the ear is enormous, particularly for notes of medium frequency where the greatest intensity that can be tolerated (the threshold of pain) is some 10^{13} times that corresponding to the threshold of hearing, for which the R.M.S. pressure is a fraction of a millidyne per cm^2. For very low and high notes the thresholds of hearing are associated with substantially higher intensities.

It is obviously more convenient to use a geometrical rather than an arithmetical scale to deal with such large ranges of intensity, and the unit adopted for the purpose is either the *bel*, which signifies a 10-fold increase in intensity or energy, or, more usually the *decibel* (db.), which signifies approximately a 5/4 increase (i.e. antilog 0·1). Two decibels denote approximately a $(5/4)^2$ increase, three decibels a $(5/4)^3$ increase, and so on, ... 10 decibels a $(5/4)^{10}$, i.e. a 10-fold increase, i.e. a bel.* Thus the upper auditory limit of intensity for a frequency of 1000 c./sec.

* Two similar sounds of intensities I and I_0 are said to differ in intensity by n decibels, where $n = 10 \log_{10} (I/I_0)$.

can be expressed either as 13 bels or 130 db. above a zero intensity corresponding to the threshold of hearing. Incidentally a decibel corresponds roughly to the smallest change of intensity which can be detected by the ear in the case of sounds of medium frequency and moderate loudness.

In most cases when sound is generated, the sound is only a small by-product, the amount of energy so expended being trifling. For instance, in hand clapping or in a collision only about 1 part in 1000 of the energy dissipated goes into sound. The figure is increased to a few per cent in the case of motor horns where the acoustic output ranges from about 1/10 to 1 W. The loud-speakers used for talking pictures may have an efficiency of 30 per cent or more, with an output of a few watts. The power of the human voice is very small, no more than about 10 μW. during normal conversation, about 100 μW. when talking loudly, and about 1000 μW. when shouting. At the other end of the scale are ship's sirens and fog sirens which attain figures in the region of a horse power.

The measurement of acoustic energy is commonly effected by an intensity meter consisting of a microphone and amplifier together with a suitable rectifier and indicator. The type of microphone used depends on the conditions. Moving coil or ribbon microphones are sometimes employed, though for fundamental work the condenser or the crystal microphone is preferred despite the lower sensitivity. The response curves of Rochelle-salt microphones are extremely uniform, and it is not without interest that the piezo-effect which has been known for more than 150 years, should have found another sphere of activity.

§ 5. ABSOLUTE SOUND-MEASUREMENTS

Whatever the microphone used, it is usually calibrated by interchange comparison under appropriate conditions with standard (pressure) microphones of the condenser type which have received fundamental calibration in absolute units by means either of the Rayleigh disc or the pistonphone. The Rayleigh disc depends for its operation on the tendency of a small thin glass disc suspended from its edge by a fine fibre to set itself at right angles to a sound field. It measures the sound particle velocity which is determined from the measured torsional constants of the system and the deflection produced. The corresponding sound pressures are deduced from the known relations in the field. The pistonphone, which measures sound pressures directly, consists essentially of a small cavity, one face of which is closed by the diaphragm of the standard microphone, the opposite face consisting of a small piston connected to the moving coil of a loud-speaker unit. The amplitude of motion of the piston when it is set in vibration is measured optically and enables the corresponding sound pressures in the cavity to be calculated.

The pressure sensitivity of the standard microphone (i.e. the ratio of the response of the microphone to the sound pressure actuating the diaphragm) is determined by the pistonphone for frequencies from 10 to about 400 c./sec. For higher frequencies (about 60–3500 c./sec.) the pressure sensitivity is determined from Rayleigh disc measurements in stationary waves in a pipe, one end of which is closed by the microphone while at the open end is mounted a loud-speaker.

The field sensitivity of the standard microphone (i.e. the ratio of the response of the microphone to the sound pressure in a progressive wave) is determined by the Rayleigh disc in a highly absorbent room for frequencies from about 300 to 10,000 c./sec. At lower frequencies, even with rooms heavily lined with cotton wool, the residual reflections from the walls are no longer negligible, so that the relation between sound pressure and particle velocity ceases to be valid.

At frequencies below about 300 or 400 c./sec. the pressure and field sensitivities of the standard microphones are indistinguishable, but as the frequencies are raised the field sensitivity becomes greater than the pressure sensitivity.

In the case of velocity microphones, which respond wholly or in part to the pressure gradient, special calibrating arrangements are necessary for field sensitivity determinations at low frequencies.

§ 6. THE ANALYSIS OF SOUND AND NOISE

The wave-form of the microphonic current of a sound can be readily displayed by means of a cathode-ray oscillograph, but the analysis of such wave-forms is usually troublesome, and the composition of a sound or noise is effected more conveniently by other means. The methods of spectrum analysis in optics are not directly applicable to the case of sound by reason of the enormous difference between the wave-lengths in the two cases and because of the fact that, whereas the frequencies of the visible spectrum cover a range of only about one octave, the range of frequencies of sound to which the ear responds extends over about ten octaves.

For the analysis of sound, the earlier experimenters used acoustical resonators which were tuned to respond to the different components present in the complex sound. In present-day methods the oscillatory pressure variations which constitute the sound are first converted, by means of a microphone, into corresponding electrical oscillations; and the problem is reduced to the analysis of a complex electrical oscillation, a process which may be effected by tuning or filtering circuits or by the use of a search or heterodyne tone. Such filters may be of the high, low or band-pass type.

Perhaps the simplest method for sound analysis is the one directly analogous to the purely acoustical method, that is, the use of electrical resonant circuits which can be tuned to respond in turn to the different components of the complex oscillation. An elaborate apparatus, completely automatic in operation, has been designed on these principles by Wegel and Moore, but the method suffers however from the disadvantage that it is not easy to cover the whole audio-frequency range continuously and quickly, nor to obtain the same selectivity or resolving power at all parts of the range. Moreover the best selectivity which can be obtained with simple resonant electric circuits is not always adequate, when closely spaced components are present.

A similar method makes use of a series of electrical band-pass filters to determine the energy in successive bands of frequencies of perhaps a half or one octave in

width. This method is particularly valuable where the sound to be analysed is not of a steady nature containing discrete components—traffic noise for instance.

An improved method which gives a constant and high degree of selectivity over a wide range of audible frequencies was devised independently, and in slightly different forms, by Grützmacher and by Moore and Curtis. This method, the "search tone" method, enables the frequency range to be covered continuously and comparatively rapidly, and is now widely used. As before, the complex sound or noise vibrations are first converted into the corresponding electrical vibrations. Upon this mixture of frequencies is superimposed an electrical oscillation of pure sine-wave form and continuously variable frequency—the search tone—in such a manner that combination tones are formed. As the frequency of the search tone is varied, the frequencies of the difference tones formed between the search tone and the successive components of the mixture pass in turn through some selected value. A sharply tuned detector arranged to respond to this selected frequency reveals each component.

In Meyer's form of the apparatus, the search tone is varied from 20 cycles upwards and the selected frequency at which the successive difference tones are detected is approximately zero. This is achieved by passing the combination tones through an electrical filter designed to pass only frequencies less than 20 c./sec. As the search tone comes within 20 c./sec. of a component of the mixture, a difference tone is formed which can be passed by the filter and recorded. Since the apparatus will not respond to zero frequency a double response is obtained for each component.

In the form of the apparatus devised by Moore and Curtis the search tone is varied from 11,000 to 16,000 c./sec. and the selected frequency at which the successive difference tones are detected is 11,000 c./sec. The detector in this case is a very sharply resonant steel bar which is set into vibration electrically each time the search tone reaches a frequency of 11,000 c./sec. greater than that of a component of the mixture. The vibration of the bar is detected electrically and recorded.

In some commercial forms of this apparatus a quartz crystal responding to a frequency of 50,000 c./sec. is used as the detector, the search tone being varied from 50,000 to about 60,000 c./sec. A high degree of selectivity is thus obtained enabling closely spaced components to be completely separated.

Recently Meyer and Thienhaus have developed an acoustic spectroscope which combines some of the features of the search tone method and of the optical method of producing a spectrum. The electrical mixture from the microphone is used to modulate a carrier frequency of 45,000 c./sec. in a similar manner to that used in a radio-transmitter. The lower side band of the modulated carrier is suppressed and the upper side band containing frequencies equal to 45,000 plus those of the modulating mixture is reconverted into sound—or rather into supersonic radiation—by means of a ribbon loud-speaker. The wave-lengths of the original sound are thus reduced some 100-fold and the original wide range of frequencies is reduced to considerably less than an octave. A spectrum of this supersonic radiation is then

obtained in a manner exactly analogous to that used in the optical case—by means of a concave diffraction grating. The grating consists of a row of some 300 parallel steel needles 3·4 mm. in diameter and 1 cm. apart, whereas for the normal acoustic spectrum a grating the size of a garden fence would be required. The first order spectrum is detected by traversing across it a microphone, specially designed to respond to these high frequencies, and connected to an amplifier and recording gear. Meyer and Thienhaus claim for their method the advantage that the spectrum can be recorded in a fraction of a second only, enabling sounds of quite short duration and variable intensity to be analysed.

Steady sounds or noises readily lend themselves to analysis and yield line spectra, but the matter is less simple with transients, that is, acoustical impulses which start suddenly and die away rapidly. Examples of transients are speech (and in particular certain consonants like *p* and *b*), whip cracks, rifle shots, blows, foot-steps, hand clapping, percussion instruments, e.g. the piano, drum, and most musical instruments. All such sounds are found to be largely of the band spectrum type with superposed line spectra and nearly all contain very high frequencies.

Sound analysis may be of particular service in studying machine noises as a means of tracking the source of major components. Synthesis of sounds is sometimes resorted to for particular purposes.

§ 7. THE MEASUREMENT OF LOUDNESS

Of the subjective features of sounds and noises, pitch and timbre have their greatest significance in music. Certain workers have attempted to evaluate the nuisance or annoyance factors of sounds but the results do not find general acceptance. In very many cases experience shows that annoyance is bound up with sheer loudness, almost the only subjective feature of a sound for which subjective standards have been attempted. In this connexion we may recall some of the peculiarities of the ear.

Figure 1 shows the auditory diagram of the average ear under free air conditions as determined by Fletcher and Munson in 1933. Loudness contours are plotted for pure tones of different frequencies and for different intensity levels measured in decibels above a specified zero. Reference to figure 1 brings out a number of very interesting features. We see, for example, that at moderate intensities the ear is more sensitive to medium frequencies than to equally intense high or low frequencies. At feeble intensities the selectivity of the ear is still more pronounced, the sensitivity being much greater for medium frequencies than for equally intense high frequencies, and very much greater than for equally intense low frequencies. Only at high intensities do equal increments of energy produce approximately equal increments in loudness at different levels.

Furthermore the thresholds of hearing are very different for different frequencies, so that the range of intensity with which the ear can deal is more restricted for high and particularly for low notes than for notes of medium frequency. Moreover, in the case of complex sounds or noises, the loudness depends upon

their character and in general we cannot calculate the loudness from the known energy-loudness relations of the pure-tone components of a complex sound.

It is evident that while loudness is clearly related to energy, the two do not always keep in step, especially for low and high frequencies, so that an energy meter, such as a microphone system, cannot unreservedly be used as a direct measurer of loudness.

The situation is met by setting up a standard reference scale of loudness obtained by selecting a pure tone of specified frequency and operating it by a progressive series of measured steps of energy. On the British Standard Scale of loudness the reference tone has a frequency of 1000 c./sec. and an *arbitrary zero*

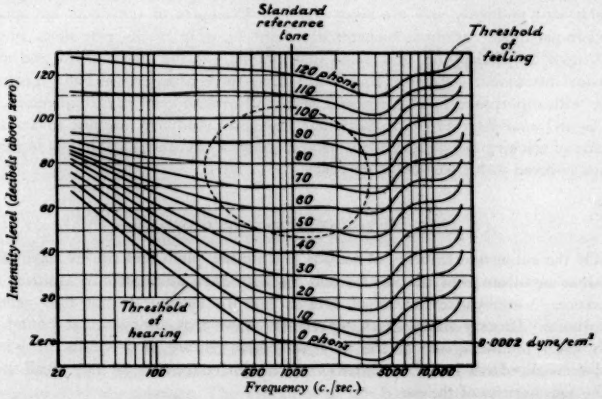


Figure 1. Auditory diagram of ear showing loudness contours for different frequencies. Free-air conditions. Fletcher and Munson, 1933.

of energy near the threshold of hearing. If the reference tone is operated by successively increasing decibel steps of energy, the resulting changes of loudness are expressed in numerically identical steps on a scale of *phons*. That is if the reference tone is excited by an intensity of n decibels above the zero, the loudness is n phons. The *equivalent loudness* of any other sound or noise is evaluated by matching it by ear against the suitably adjusted reference tone, the numerical value of the latter in phons then giving the equivalent loudness of the sound to be measured. By this procedure, which rests on the admitted ability of the average individual to match equality of loudness, we set up a subjective scale of equivalent loudness, the unit being the phon.

The experimental realization of the fundamental scale of phons is one for the standardizing laboratory, as will be appreciated from the following definition of the phon taken from the *British Standard Glossary of Acoustical Terms and*

Definitions, 1936 (No. 661). A closer specification of the experimental conditions is at present being developed by the British Standards Institution.

"Phon (B.S.) A unit of equivalent loudness, defined as follows. The standard tone shall be a plane sinusoidal sound wave train coming from a position directly in front of the observer and having a frequency of 1000 cycles per second. The listening shall be done with both ears, the standard tone and the sound under measurement being heard alternately and the standard tone being adjusted until it is judged by a normal observer to be as loud as the sound under measurement. The intensity level of the standard tone shall be measured in the free progressive wave. The reference level shall be taken to be that corresponding to an R.M.S. sound pressure of 0·0002 dyne per sq. cm. When, under the above conditions, the intensity level of the standard tone is n decibels above the stated reference intensity, the sound under measurement is said to have an equivalent loudness of n phons (B.S.)."

Table 1 sets out the B.S. scale of loudness in simple form.

Table 1. B.S. Scale of equivalent loudness. Standard tone: 1000 c./sec.; plane wave in free air. Zero: 0·0002 dyne/cm^2, near threshold of audibility. Listening with both ears, observer facing source

Intensity of standard tone		Loudness of standard tone in phons	Popular description
In terms of the zero intensity	In decibels above zero		
1	0	0	Silence
10	10	10	
10^2	20	20	Quiet
10^3	30	30	
10^4	40	40	
10^5	50	50	Moderate
10^6	60	60	
10^7	70	70	
10^8	80	80	Loud
10^9	90	90	
10^{10}	100	100	
10^{11}	110	110	Very loud
10^{12}	120	120	
10^{13}	130	130	

Various zeros of loudness have been employed in the past, e.g. 1 millidyne/cm^2 which results in numerical values of loudness some 10 phons less than with the B.S. zero. In Germany the phon is based on a zero of 0·0003 dyne/cm^2 which is equivalent to a 4 db. difference from the B.S. zero. As however a different method of listening is employed the exact discrepancy between the two scales is not known. The American scale agrees with the British, except that in the States it is customary to use the decibel not only for expressing intensity measurements, but also for loudness levels, it being implied in the latter case that the decibel figure quoted refers to the energy level of the standard tone when it matches the sound to be measured.

Incidentally a glance at figure 1 shows that in the centre of the diagram (approximately in the region of the dotted circle) the phon contours are roughly horizontal and parallel, so that over a wide range of pure tones from, say, 200 or 300 to 3000 or 4000 cycles, and of intensities from, say, 40 or 50 db. to 90 or 100 db., the loudness level in phons is approximately equivalent numerically to the intensity level in decibels. As many everyday noises are largely made up of components within these ranges and have their maximum level between 500 and 1000 cycles, a similar relation may hold sufficiently well for such noises if there is not undue intermasking between the components.

Reference may here be made to the criticism to which the phon scale has been subjected in some quarters in that it does not interpret numerically one's sensations of loudness, a phon near the top of the scale admittedly signifying a much bigger step than it does near the bottom. Some workers* have accordingly attempted to set up a subjective scale of loudness units in which the assumption is made that fractional estimation is possible, that is, for example, it is possible to say when one sound is twice as loud as another. Proceeding in this way a general empirical relation has been derived

$$(\text{No. of loudness units}) = (\text{Loudness in phons})^{3} \times 10^{-8}$$

Table 2 illustrates the relation between the two loudness scales on the basis of a common reading of 100. As will be seen, they have little in common numerically. The scale of loudness units is however based on assumptions which

Table 2. Comparison of subjective scales

Loudness of standard tone	
Phons, B.S.	Fractional estimation (loudness units)
0	0
10	0·05
20	0·15
30	0·4
40	1·2
50	3
60	7·5
70	17
80	32
90	60
100	100
110	160
120	250
130	370

are not universally recognized, and while it is conceded that the phon scale does not provide a linear scale of sensation, and that the region below about 40 phons is of little practical significance, we have to ask ourselves whether such objections really matter? We do not, for example, expect water at 90° F. to feel twice as hot

* Richardson and Ross (1930); Ham and Parkinson (1932); Geiger and Firestone (1933); Churcher, King and Davies (1934).

as at 45° F., neither need we expect 90 phons to sound twice as loud as 45. Many of us buy sheet metal, screws, nails, wire, shoes, buttons, bacon, etc., and find no particular inconvenience in the apparently irrational scales employed. It is obviously a matter of experience, and while a scale of equivalent loudness in phons may not be subjectively ideal, it does rest on sound physical foundations and accepted subjective principles. Furthermore the close association of the phon scale with the decibel scale is of great practical value to the acoustical engineer, for example in dealing with the transmission of air-borne sounds by walls.

Noise meters. The conditions laid down for the fundamental realization of the phon are not those which are likely to obtain in everyday circumstances of noise measurement, and in practice simpler techniques are employed. There are two main types of commercially available sound and noise meters on the market, namely subjective meters of the Barkhausen type, and objective meters of the microphone type. The calibration of such meters under the particular conditions of use by means of absolute standards is work which demands the resources of a standardizing laboratory. The British Standards Institution is at present engaged in developing suitable secondary methods, but at present there is not enough experience to say definitely whether the subjective standard of measurement can be satisfactorily realized in everyday practice.

§ 8. SUBJECTIVE METERS

Subjective noise meters depend on the equality matching of the loudness of a noise as heard by the ear with a reference tone (usually a pure tone of specified frequency and of graduated intensity) as heard in a telephone earpiece held tightly against one ear. The reference tone may be produced by an electric buzzer, valve oscillator or other means, various frequencies (e.g. 800 or 1000 c./sec.) or mixed tones being used in commercial instruments. The usual Barkhausen technique involves simultaneous listening of the noise and the reference tone, but experience indicates that alternate listening is preferable when it can be arranged.

Subjective meters are normally graduated in phons, and their calibration in terms of subjective readings is carried out in accordance with the definition of the B.S. phon by the measurement of laboratory controlled noises.

Subjective meters are useful for certain purposes, but the aural judging of equality of loudness of the reference tone and of a noise very different in character is not always easy. Individuals are found to differ widely in their judgements and the same individual is not always consistent. The accuracy of judgement is normally low, and a team of experienced observers is essential for more precise work. The difficulties are emphasized in the case of unexpected or short-lived impulsive sounds.

§ 9. OBJECTIVE METERS

The design of objective sound and noise meters has received considerable attention in the United States, Germany and this country. All the various objective meters consist essentially of a pressure microphone connected to an amplifier

provided with calibrated gain control, followed by some type of rectifier and an indicator graduated in phons or decibels. The amplifiers contain one or more electrical networks designed to modify the shape of the frequency characteristic so that it imitates the ear sensitivity at selected loudness levels.

In this country and Germany the aim has been to design objective meters which will measure sounds and noises on the subjective scale of phons, that is to simulate the selectiveness and response of the average ear in different circumstances. The meters are first of all constructed to give the same reading for steady pure tones of different pitches which sound equally loud to the ear. In the case of intermittent sounds, whether isolated impulses or recurrent at a variety of rates, not all objective meters are so designed to simulate the salient characteristics of the ear in integrating such sounds. For example, the ear does not record full strength until a steady sound has persisted for about $\frac{1}{5}$ sec., and in the case of pulsating sounds the inertia of the instrument has to be of the right order to give readings corresponding to average aural appraisement.

In the United States the present tendency is to design objective meters, which will stand, so to speak, on their own legs. The meter is constructed to a standard arbitrary specification and is designed to read approximately correctly for a range of pure tones. The readings are referred to as "sound levels" and no claim is made that they conform to subjective "equivalent loudnesses".

The experience of the National Physical Laboratory is that while objective meters, whether of the noise or sound level type, may give reasonable agreement with subjective measurements in the case of pure tones or other steady sounds, such meters will, unless properly designed, give very low readings in the case of intermittent noises, such as traffic noise. So far, no standard specification has been introduced into this country for objective meters, and if it is considered that the primary requirement of such a meter is that its readings should conform reasonably well with subjective measurements carried out in accordance with the definition of the British Standards phon, then it would appear that standardization at the present juncture is premature and liable to lead to a multiplicity of loudness scales calculated to confuse the public. In the meantime, an instrument has been designed at the National Physical Laboratory which satisfactorily conforms with subjective measurements for both steady and intermittent noises at moderately high loudness levels. An account of this will shortly be published.

There is no doubt that the future of sound or noise measurement from a practical point of view lies with the objective meter. A dial instrument is always preferred by the untrained observer, and even in the present state of development the practical convenience of objective meters will often outweigh their theoretical imperfections.

§ 10. NOISE-MEASUREMENTS

For some years the measurement of noise and the investigation and analysis of noises of varied origin, with a view to their suppression, has formed a considerable proportion of the acoustical work of the National Physical Laboratory. Some of these noise levels are shown in figure 2. We may take it that in most cases

the abatement of noise levels above 70 phons would be in the interests of the community.

Since 1929 an extensive programme of work has been carried out for the Aeronautical Research Committee of the Air Ministry on the various aspects of aircraft noise and its abatement. Measurements have been made in aeroplanes in flight, and on the noise from airscrews and engine exhausts under experimental conditions. In connexion with the development of exhaust silencers an investigation has been carried out on model silencing conduits considered as acoustic filters.

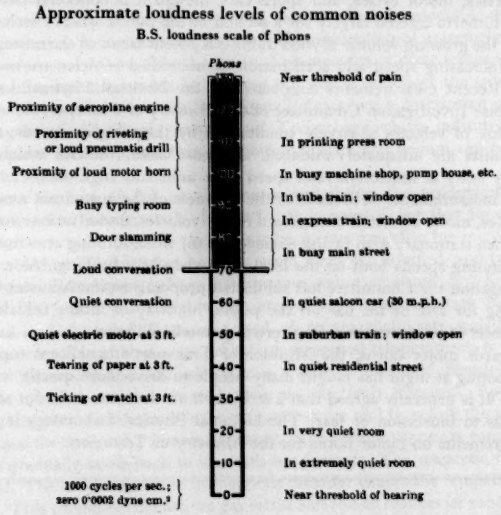


Figure 2.

The Laboratory has also carried out an investigation of train noises for one of the railway companies, and is frequently consulted in connexion with the suppression of noise in such situations as aeroplane engine testing beds, printing works, transformer substations, noisy offices and so on.

Road transport noises are much in the public mind at the moment and are at present the subject of enquiry by the Ministry of Transport. For vehicles fitted with pneumatic tyres there does not seem to be much to choose, from a noise point of view, between present-day road surfaces of good quality though concrete roads are noticeably louder than tar macadam. The major components of motor vehicle noises are those from the exhaust, engine and transmission, though from the point of view of the public it is the over-all noise which matters. The motor industry

as a whole deserves great praise for its efforts to secure the quietness of motor vehicles. This applies particularly to private cars, most of which when normally used are unobjectionable from the point of view of noise. The flexible mounting of engines, the counterbalanced crankshaft, silent gearing, the straight-through type of exhaust silencer are all major contributions. In the matter of exhaust silencing certain motor car constructors have achieved remarkable perfection. The newer designs of buses are attractively quiet, and the replacement of the tramcar by the trolley bus has been an acoustic boon. Complaints are however still frequent of certain lorries, motor cycles, and sports cars, though it is conceded that the last two have hitherto catered largely for a section of the public which revels in noise.

While the growing volume of road traffic is a potent factor in the noise problem, the ever increasing speed and acceleration of individual vehicles are no less important. Recent measurements conducted by the National Physical Laboratory for the Noise Investigation Committee of the Ministry of Transport show that the overall noise of vehicles is largely conditioned by their engine speeds, that is if their exhausts are adequately silenced. In most cases, vehicles which are exceptionally noisy at maximum road speed have unusually high maximum speeds. Objective measurements were made of the loudness of the noise from a number of motor cycles, motor cars, and commercial motor vehicles, under various conditions, e.g. (*a*) when stationary with racing engine, and (*b*) when running at various steady and accelerating speeds both on the level, and when hill climbing. As a result of the investigation the Committee has submitted proposals to the Minister to forbid the offering for sale or for use on the public highway of motor vehicles which produce noise of loudness exceeding certain prescribed limits.

As regards motor horns, the Minister of Transport's beneficent suppression of horn hooting at night has taught many people to drive more quietly in the day time also. It is generally agreed that a strident horn does not make for safety but rather leads to indecision or fear. The National Physical Laboratory is carrying out measurements on motor horns for the Ministry of Transport.

HEAT

By R. W. POWELL, B.Sc. Physics Department,
The National Physical Laboratory, Teddington, Middlesex

§1. PRODUCTION AND MEASUREMENT OF VARIOUS TEMPERATURES

One of the aspects of the subject of heat to which considerable attention is at present being directed is that of the production of extremely low temperatures. At one time the study of phenomena at these temperatures was largely confined to the Leiden Laboratory but this is no longer the case. Several cryogenic laboratories now exist, and Goetz[1] has given a description of the new laboratory at the California Institute of Technology which indicates that the essential plant for the production of liquid hydrogen and helium can be installed at a moderate cost.

A clear account of the physical basis underlying the production of these temperatures has been given by Simon[2]. It is pointed out that for a lowering of temperature to be possible there must be means available for increased order to be established. This is what happens when gases are liquified, and by reducing the pressure over liquid helium a minimum value of about 0·7° K. has been obtained. To attain still lower temperatures the demagnetization method first suggested by Debye and Giauque is used. A paramagnetic salt, sealed in a small glass tube containing about 1 cm. of helium at normal temperatures, is first cooled in liquid helium. Within the salt there is a state of magnetic disorder, since the magnetic moments attached to each atom are randomly disposed. Application of a strong magnetic field increases the magnetic order and therefore the temperature rises, since the total order remains constant. The helium vapour is conducting and the salt gradually cools back to the bath temperature. The magnetic field is then removed. The sudden increase in magnetic disorder causes the temperature to fall rapidly. This condenses the helium gas on the salt and so reduces its vapour pressure that the salt becomes insulated from the enclosure. In consequence of this low pressure and of the fact that radiation is negligible, a temperature of 0·3° K. or less once attained is very easily maintained.

Giauque and MacDougall[3] starting from 1·29° K. and using gadolinium sulphate with a field of 8000 G. have reached a temperature of 0·242° K.; Kürti and Simon[4] using iron ammonium alum and a field of 14,000 G. reached 0·038° K., whilst de Haas and Wiersma[5] using a diluted chromic alum and a field of 24,075 G. have reached 0·0044° K.

The foregoing temperatures have been calculated from measurements of the magnetic susceptibility, on the assumption that the salt obeys Curie's law. It is important to ascertain how this scale agrees with the thermodynamic temperature scale. Kürti and Simon[6] have pointed out that the entropy of the salt in zero field

can readily be obtained as a function of the Curie temperature scale. If, in addition, calorimetric measurements are carried out the true thermodynamic scale can be established, as, according to the second law of thermodynamics, the absolute temperature can be calculated once the changes of heat content and entropy between two states are known. Preliminary measurements using iron ammonium alum indicate that above 0·10° K. the two scales are in agreement to within the 2 per cent error involved in the experiments. At a Curie temperature of 0·06° the absolute temperature has about a 10 per cent higher value. In a preliminary note Giauque and MacDougall[7] have stated that when gadolinium phospho-molybdate is used, at 0·3° K. the Curie scale is within 0·01° of the true thermodynamic scale, at 0·25° K. it is 5 per cent low, whilst at 0·15° K. it is 22 per cent low. A departure in the opposite direction had been anticipated, and further information on this matter will be awaited with interest.

Above say, 1·2° K., vapour pressure or gas thermometers are used as standard temperature measuring instruments. They are not, however, always most convenient, and recourse has recently been had to resistance thermometers for direct measurements at temperatures below 10° K. Pure metals have a very small temperature coefficient in this range and most supra-conductors change too rapidly at the transition temperature to be employed for this purpose. Leiden workers had reported that the resistance of phosphor-bronze shows a gradual decrease below 7° K. Babbitt and Mendelssohn[8] have accordingly examined over a dozen different phosphor-bronze specimens, one of which proved suitable for use as a thermometer down to 2° K. An alloy of silver with 5 per cent lead proved suitable from 7 to 3° K., and it is hoped eventually to find a resistance material which can be used over the entire helium range.

At temperatures ranging from 0 to −183° C. Keesom and Dammers[9] have compared five platinum thermometers, which satisfied the conditions defined in the international temperature scale, with the helium thermometer. They conclude that between 0 and −100° C. the international temperature scale lies below the thermodynamic scale, differing by 0·04 at −80° C., whilst between −120 and −183° C. it lies above, the maximum deviation being between −130 and −140° C. In this work the oxygen point had been accepted as $-182{\cdot}98_3$° C., in accordance with a new determination.

For the measurement of temperatures ranging from about 12 to 2000° K. thermocouples composed of various elements are being increasingly used, and a number of papers have been published in this connexion during the past year. Aston, Willihnganz and Messerly[10] compared the electromotive forces of nine copper constantan thermocouples against a helium thermometer. The measurements were made from 12° K. to the ice-point, which was taken as 273·16° K., and the results are tabulated from 12 to 90° K. These supplement the values obtained in 1929 by Southard and Andrew who calibrated a copper-constantan thermocouple against a resistance thermometer down to 85° K., at which temperature the two calibrations differ only by 0·9°. It does not, however, appear that the composition of all constantan wires is sufficiently constant for this calibration to be used without

checking, since Aoyama and Kanda[11], who also calibrated a copper-constantan thermocouple against a gas thermometer down to a temperature of about 80° K., obtain results differing by some 20° at these low temperatures. Below 20° K. the thermocouples become very insensitive.

In this connexion it is interesting to note that the application of vacuum tube amplifiers for the magnification of thermocouple electromotive forces has formed the subject of two papers[12,13]. Razek suggests a simple means for combining the small direct voltage with an alternating carrier. The two are then passed through a copper oxide rectifier which suppresses one part of the cycle to an extent depending on the magnitude of the direct voltage. The resulting fluctuating voltage can be amplified to any desired degree.

With regard to the other end of the temperature scale Keinath[14] has reviewed some of the combinations of non-metallic elements which are available for use at high temperatures and directed attention towards the carbon-silicon carbide thermocouple introduced a few years ago by Fitterer. This combination gives an electromotive force of about 300 μV. per ° C., and the values are stated to be reproducible to within about 1 per cent up to some 1700° C.

Workers[15] at the National Bureau of Standards have published standard tables for the calibration of chromel-alumel thermocouples, composed of wires supplied by the Hoskins Manufacturing Company, over the temperature range from 0 to 1400° C.

A full description has also been given[16] of the methods employed at this institution when calibrating thermocouples. The best methods of interpolating for temperatures between the standard temperatures are also discussed and a table giving the accuracies obtainable with various types of thermocouples is included.

A comprehensive paper dealing mainly with various sources of error which may arise in the practical use of thermocouples has been published by Euler and Guthmann[17]. Not only does this paper deal with methods of use which avoid errors due to conduction from the hot junction, but also with the auxiliary equipment, and the various types of protecting tubes required to prevent contamination. Pythagoras tubing is considered to remain gas tight up to 1580° C., whilst Siemens' Sinter-korund is gas tight up to 1720° C. For use above 2000° C. oxides of beryllium, thorium and zirconium are available.

It is apparent from the aforementioned papers, and also from the Report of the National Physical Laboratory for 1935[18], that steadily increasing precision is being obtained in thermocouple measurements. At the last-mentioned laboratory a thermostatically controlled block of nickel is now used as a bath to reduce the temperature gradients in the ingot when a melting- or freezing-point determination is being made. This ensures a sharply defined temperature arrest, and enables the silver point to be reproduced to 0·01° C. and the antimony point to 0·03° C.

Although the melting points of the metals of the calcium family are never likely to serve as even secondary standard points, their recent determination by Hoffmann and Schulze[19] is of interest in that the new values differ considerably from those accepted at the time of publication of the *International Critical Tables*. The new

determinations were made on the sample heated in an enclosure filled with pure argon so as to prevent the combination with nitrogen which may have accounted for the divergent earlier results. The values obtained are given in table 1.

Table 1

Metal	Melting point given by		
	Hoffmann and Schulze	Rinck (1931)	*I.C.T.*
Calcium	851	848	810
Strontium	771	—	800
Barium	704	710	850

It is seen that Hoffmann and Schulze's values are supported by the results of Rinck, and that the melting point now decreases as the atomic volume increases, as has been found for elements of similar groups of the periodic table.

A re-determination of the melting point of palladium by the ratio of brightness method has been carried out at the National Physical Laboratory[20]. For this measurement an inductively heated ingot was used and the technique was similar to that recently employed in the case of platinum. The value obtained is 1554·4 ± 1·0° C. and it is satisfactory that this result confirms the earlier data which led to the recommendation of the value of 1555° C. for the international temperature scale.

The actual measurements at these high temperatures were, of course, made by means of a disappearing filament optical pyrometer operated under ideal black body conditions. To obtain the true temperature of a surface which is radiating freely to its surroundings it is necessary to know the emissivity of the surface for the wave-length used. The emissivities of columbium, thorium, rhodium and molybdenum have been measured by Whitney[21], and for $\lambda = 0{\cdot}667\,\mu$ found to be 0·374, 0·380, 0·242 and 0·382 respectively over the approximate temperature ranges of 1300–2100° K. The metals were in the form of a V-wedge or of rolled cylinders and were heated in a vacuum of at least 10^{-6} mm. of mercury.

For the past twenty years pyrometers of the disappearing filament type have been constructed in accordance with certain restrictions arising from the experimental investigations of Fairchild and Hoover (1912) and of Worthing and Forsythe (1914). These were suggested in order to minimize the brightening of the edge of a round filament and the dark outer line. Cunnold[22] has recently demonstrated that the effect of any reflexion of light from the source at the surface of the filament, such as was suggested by Fairchild and Hoover, should be quite negligible. The bright line within the edge of the filament is definitely attributed to deviations from Lambert's cosine law of emission. This conclusion was supported by photometric examination of photographs of a filament. The diffraction effects which can occur in an optical pyrometer are also considered, and are shown to result from the obstruction of portions of the Airy patterns in the filament plane. They must always be present, but can be rendered indiscernible by suitable adjustment of the relative

sizes of the entrance and exit apertures. In the past it has generally been assumed that perfect disappearance of the filament is necessary as, otherwise, the brightness ratio may differ from unity, and further, the presence of a dividing line between the filament and the image of the source might lead to a decrease of accuracy for a single observer and to inconsistent results for a number of observers. Cunnold, however, realized that so long as the brightness ratio remains constant it need not be unity, and actual tests were made by a number of observers on a new pyrometer in which the fixed exit stop had been replaced by an iris diaphragm so that the exit aperture could be varied from 0·02 to 0·07 radian. The results show a definite increase of accuracy with increase of exit aperture, although the disappearance became less perfect. It has thus been experimentally established that perfect disappearance is not really essential. This result has important applications in the design of pyrometers, for it allows the restrictions previously imposed to be considerably widened and enables pyrometers transmitting much more light to be constructed. This is of twofold importance, since not only can more truly monochromatic filters be used, but the optical scale can be extended downwards, and this may render it possible to obtain a value of Wien's constant, C_2, in terms of the optical scale.

The disappearing filament pyrometer described by Cunnold and his colleagues[23] has an exit aperture of 0·0628 radian, magnifications of 20 and 5 with corresponding working distances of 24 and 48 cm. By measuring the filament current to 0·01 per cent, temperatures should be measurable to 1 in 3000 at the gold point.

An optical pyrometer departing from the normal type[24] has been marketed by Messrs Elliot Bros. In this instrument the filament is maintained at a constant temperature and the brightness of the source adjusted by interposing a wedge-like graduated absorption screen. The novel feature consists in the use of two crossed filaments composed of materials having quite different temperature-current characteristics. The two are equally bright at one temperature only. By setting the pyrometer for this condition the necessity for an instrument to measure the filament current is avoided.

When conditions are favourable the mercury in glass type of thermometer has advantages, and Moser[25] has found that by using as thermometric liquid an eutectic alloy of 91·5 per cent mercury and 8·5 per cent thallium an expansion thermometer is obtained which can be used down to −60° C. Moser[26] has also described an improved method of filling the space above the mercury column with an inert pure gas exerting a pressure exceeding that of the mercury to enable the range of the thermometer to be extended. The method involves the condensation of the requisite quantity of argon on the mercury by cooling the thermometer with liquid air. A litre of specially purified argon at atmospheric pressure is sufficient to fill several hundred thermometers, and in addition to being simpler in application, the method is cheaper than that previously used, which involved a large pressure chamber and some thousands of times the quantity of purified nitrogen.

A somewhat novel type of thermometer which is now available for use over the range −200–1000° C. is one in which advantage is taken of the variation of the

refractive index of vitreous silica with temperature. Measurements of this property having been made over the above range[27].

With regard to the measurement of a particular class of temperatures that of liquid steel is a matter of considerable importance, and one for which an entirely satisfactory method has not yet been found. The Heterogeneity of Steel Ingots Committee has been dealing with this subject. In their last report[28] a disappearing filament pyrometer is recommended for measurements made during pouring. Measurements on the liquid when in the furnace are more difficult, and it is stated that tests are being made using a thermocouple of the Fitterer type, to which reference has already been made. Other methods which have been suggested[29] are: a thermocouple of the platinum type to be immersed in the bath for a brief time only, special precautions being taken to prevent deterioration before and after immersion; the immersion of small samples of alloys of known melting point in the bath and examination to see which have melted; sighting an optical pyrometer on to a "black body" immersed in the bath, e.g. the interior of a tube dipping into the molten steel, or a carbon target immersed in the steel and momentarily lifted out while observations are taken. The practicability of the first of these depends on the production of improved sheathing materials, and provided the effect of fumes can be overcome the method last suggested, at the moment, appears to be the most promising.

Attention should be directed to a paper by Blaurock[30] dealing with the measurement of the temperature of liquid iron and steel by optical methods. After trying out several methods the best was considered to be one in which measurements were made using first a red and secondly a green filter in the pyrometer, independent experiments having shown that the points fall approximately on a straight line if the true temperature minus the apparent temperature with the red filter is plotted against the difference between the two apparent temperatures using the red and green filters. By using this line to derive the true temperature from the two observed values it is considered that the temperature of the steel can be determined to 15° C. Measurements are made on both dark and bright areas of the melt.

The measurement of flame temperatures is another matter which calls for comment. The spectral line reversal method has frequently been used for this purpose. Following the criticism of this by David, Lewis and von Elbe have supported the reversal method and have pointed out the possibility that David's own resistance thermometer method of measuring flame temperatures may not be accurate. There have been several more controversial papers[31,32,33,34,35,36] but little fresh experimental evidence, and the most satisfactory outcome would appear to be for David to adopt an alternative method such as the electrically heated wire method, so as to lend experimental support to his resistance wire data. Until additional evidence to the contrary is produced the spectral line reversal method will, no doubt, be accepted as a very convenient means of measuring flame temperatures. It has been employed in at least two instances during the past year to study flame temperatures within a petrol engine[37,38]. Rassweiler and Withrow made their measurements at three different positions within the combustion space.

Experiments made under both knocking and non-knocking conditions show that when the charge is completely inflamed, there is a temperature gradient along the length of the combustion space, the temperature at the sparking plug end being as much as 600° F. higher than at the opposite end. When knocking, the maximum temperatures were higher, and were attained earlier in the cycle, the rate of cooling during the expansion stroke was greater, and the exhaust temperatures were lower.

Determinations of arc temperatures have been made by Suits both by an optical method[39] and from sound velocity measurements[40].

The methods of thermostatically controlling temperatures are steadily improving[41,42,43]. Keesom and Dammers[9] claim to have used a cryostat which remains constant and uniform to within 0·002° C., whilst by including the windings of a platinum furnace in one arm of a bridge, the other three arms being of manganin, and allowing any out of balance deflexion of the galvanometer spot to work a photoelectric relay, a furnace can be controlled to within 0·01° C. at 1000° C.[44] Adcock[45] has described a control circuit for varying the temperature of a furnace at a pre-determined constant rate.

During the past few years a group of alternative materials to nichrome have been marketed for use as furnace windings. These include Megapyr, Kanthal, Permatherm and Alsichrom, each of which has approximately the same chemical composition (Fe 65 per cent, Cr 30 per cent and about 5 per cent Al), and a specific resistance of about $140 \times 10^{-6}\Omega$. cm. at 0° C. They can be used up to about 1350° C. compared with the 1150° C. of the nickel-chromium alloys. Hoffmann and Schulze[46] have measured the melting points of the four alloys mentioned above, using an evacuated electrically heated carbon tube furnace, and have found that each alloy melts at approximately 1500° C.

The Reichsanstalt are experimenting with tubular furnaces composed of mixtures of zirconium and yttrium oxides. These become electrical conductors when red hot, and can then be electrically heated in air to temperatures of the order of 2100° C. This temperature can be produced in a tube 1·1 cm. in diameter, 2 mm. wall thickness, and 7 cm. long by a potential of 220 V.[47]

References

(1) Goetz, A. *Rev. sci. Instrum.* **6**, 217 (1935).
(2) Simon, F. *Not. Proc. roy. Instn*, **28**, 515 (1933–5).
(3) Giauque, W. F. and MacDougall, D. P. *J. Amer. chem. Soc.* **57**, 1175 (1935).
(4) Kürti, N. and Simon, F. *Proc. roy. Soc.* **149**, 152 (1935).
(5) de Haas, W. J. and Wiersma, E. C. *Physica*, 's Grav., **2**, 335 (1935).
(6) Kürti, N. and Simon, F. *Proc. roy. Soc.* **152**, 21 (1935).
(7) Giauque, W. F. and MacDougall, D. P. *Phys. Rev.* **47**, 885 (1935).
(8) Babbitt, J. D. and Mendelssohn, K. *Phil. Mag.* **20**, 1025 (1935).
(9) Keesom, W. H. and Dammers, B. G. *Physica*, 's Grav., **2**, 1051 and 1080 (1935).
(10) Aston, J. G., Willihnganz, E. and Messerly, G. H. *J. Amer. chem. Soc.* **57**, 1642 (1935).
(11) Aoyama, S. and Kanda, E. *Bull. chem. Soc. Japan*, **10**, 472 (1935).
(12) Johnson, R. P., Bell, M. E. and Nottingham, W. B. *Phys. Rev.* **47**, 426 (1935).
(13) Razek, J. *J. Franklin Inst.* **219**, 137 (1935).

(14) KEINATH, G. *Arch. tech. Messen*, **4**, T. 55 (1935).
(15) ROESER, W. F., DAHL, A. I. and GOWENS, G. J. *Bur. Stand. J. Res.*, Wash., **14**, 239 (1935).
(16) ROESER, W. F. and WENSEL, H. T. *Bur. Stand. J. Res.*, Wash., **14**, 247 (1935).
(17) EULER, H. and GUTHMANN, K. *Arch. Eisenhüttenw.* **9**, 73 (1935–6).
(18) *Rep. nat. phys. Lab.*, Lond., p. 30 (1935). (H.M. Stationery Office.)
(19) HOFFMANN, F. and SCHULZE, A. *Phys. Z.* **36**, 453 (1935).
(20) *Rep. nat. phys. Lab.*, Lond., p. 29 (1935). See also SCHOFIELD, F. H., *Proc. roy. Soc.* **155**, 301 (1936).
(21) WHITNEY, L. V. *Phys. Rev.* **48**, 458 (1935).
(22) CUNNOLD, F. A. *Proc. roy. Soc.* **152**, 64 (1935).
(23) MILFORD, M., BRACEY, R. J., CUNNOLD, F. A. and EGERTON, A. C. *J. sci. Instrum.* **12**, 80 (1935).
(24) *J. sci. Instrum.* **12**, 303 (1935).
(25) MOSER, H. *Wiss. Abh. Phys.-tech. Reichsanst.*, Berl., **19**, pt. 5, 27 (1935).
(26) MOSER, H. *Phys. Z.* **36**, 153 (1935).
(27) AUSTIN, J. B. and PIERCE, R. H. H. Jr. *Physics*, **6**, 43 (1935).
(28) *Sixth Report. The Heterogeneity of Steel Ingots Committee.*
(29) HATFIELD, W. H. "The Work of the Heterogeneity of Steel Ingots Committee." *Iron and Steel Inst.* Special Report No. 12, p. 23.
(30) BLAUROCK, F. *Arch. Eisenhüttenw.* **8**, 517 (1934–5).
(31) DAVID, W. T. *Engineering*, **140**, 232 (1935).
(32) DAVID, W. T. *Engineering*, **139**, 195 (1935).
(33) DAVID, W. T. *Phil. Mag.* **20**, 44 (1935).
(34) LEWIS, B. and VON ELBE, G. *Engineering*, **140**, 203 (1935).
(35) LEWIS, B. and VON ELBE, G. *Engineering*, **139**, 168 (1935).
(36) LEWIS, B. and VON ELBE, G. *Phil. Mag.* **20**, 44 (1935).
(37) LLOYD EVANS, B. and WATTS, S. S. *Engineering*, **139**, 48 (1935).
(38) RASSWEILER, G. M. and WITHROW, L. *Trans. Soc. automot. Engrs*, N.Y., **30**, 125 (1935).
(39) SUITS, C. G. *Physics*, **6**, 315 (1935).
(40) SUITS, C. G. *Nat. Acad. Sci. Proc.* **21**, 48 (1935). See also *Physics*, **6**, 190 (1935).
(41) BARRELL, H. and EVANS, J. C. *J. sci. Instrum.* **12**, 281 (1935).
(42) BLOXAM, H. P. *J. sci. Instrum.* **12**, 361 (1935).
(43) WALSH, W. L. and MILAS, N. A. *Industr. Engng Chem.* (Analytical Ed.), **7**, 122 (1935).
(44) *Rep. nat. phys. Lab.*, Lond., p. 29 (1935). (H.M. Stationery Office.)
(45) ADCOCK, F. *J. sci. Instrum.* **12**, 285 (1935).
(46) HOFFMANN, F. and SCHULZE, A. *Phys. Z.* **35**, 881 (1934).
(47) TINGWALDT, C. *Phys. Z.* **36**, 627 (1935).

§ 2. CALORIMETRY AND ALLIED THERMODYNAMICS

It was mentioned in last year's report that Jessel had carried out experiments which led him to conclude that the curve of specific heat against temperature of air-free waters differs appreciably from that for air-saturated water. Laby and Hercus[48] have since criticized these experiments on the grounds that from thermodynamic reasoning in which allowance is made for (i) the heat of solution of the air, (ii) the latent heat of the water vapour required to saturate the air bubbles, and (iii) the surface energy of the bubbles, the difference in the specific heats at 15° C. of aerated and air-free water cannot exceed 1 part in 10,000, compared with about 7 parts, as found by Jessel. Jessel's result implied that the mechanical equivalent of heat should be redetermined using air-free water. Laby and Hercus proceed to show that their own and Jaeger and Steinwehr's determinations of J at

15° C. agree to within 1 part in 10,000 and that these values agree with Jessel's value for air-free water to an accuracy of 1 in 2000. They conclude that the value of the mechanical equivalent of heat for both deaerated and aerated water in which the air remains in solution is $4{\cdot}1852\times10^7$ erg. per 15° C. calorie. Jessel's result is supported, however, by an analogous result which he and Lang[49] obtained in experiments using aerated oils, and the matter appears to be of sufficient importance to warrant further investigation.

A number of direct measurements of specific heat have been carried out. These include room temperature measurements on metallic rods by Honda and Tokunaga[50] using a modification of the electrical method first employed by Nernst. The rod is wound with a heating coil and serves as its own calorimeter. The heat losses only amount to 2 or 3 per cent. A number of copper-nickel alloys are examined and a marked increase in the specific heat is observed at the composition at which the alloys become ferromagnetic.

Jaeger and his colleagues[51] have extended their measurements to include zinc, magnesium and the binary alloy $MgZn_2$, and have also determined the precautions which have to be taken to enable a metal-block calorimeter within a Dewar vacuum vessel to be used as a precision instrument[52].

Keesom and Miss Keesom[53] have measured the specific heat of liquid helium under its saturated vapour pressure. The principal change in specific heat at the lambda point occurs within a very small temperature interval, and there appears to be no latent heat connected with the transformation. It is thought that the lambda transition in liquid helium depends on interactions between neighbouring atoms.

Keesom and Clark[54] have determined the atomic heat of nickel over the range 1·1–19·0° K. It is found to considerably exceed the heat capacity due to the atomic lattice as calculated from Debye's T^3 law. From 1·1 to 9° K. the additional heat capacity expressed in calories per gram can be represented by $0{\cdot}001744T$. Mott[55] has suggested that there is a connexion between this excess heat, and the large atomic heat possessed by nickel above the Curie temperature, and concluded that in both instances the additional specific heat was due to electrons.

During 1934 researches carried out at Leiden revealed that a discontinuity occurred in the specific heats of pure metals at the temperature at which they became supraconducting. Rutgers[56] published a formula for the difference between the specific heats of the supraconducting and normal states which was in good agreement with measurements on tin and thallium. This formula indicated that the specific heat difference should be proportional to the square of the slope at the transition point of the curve obtained when the magnetic field that destroys supraconductivity is plotted against temperature. It was interesting to see what happened in the case of an alloy for which the slope is much greater than for a pure metal. Such measurements have been made by Mendelssohn and Miss Moore[57] on the alloy $PbTl_2$ over the temperature range 3–6° K., and by Schubnikow and Chotkewitsch[58] on an alloy of 65 per cent lead and 35 per cent bismuth. In neither case was a jump in specific heat of the order indicated by Rutgers' formula

found, in fact there was no appreciable discontinuity in the specific heat at the transition point. Rutgers' formula has been rigidly established by Gorter and Casimir's thermodynamical treatment[59] in which it is assumed that the transition to supraconductivity is accompanied by a disappearance of the magnetic induction. Mendelssohn and Moore suggest that a supraconducting alloy may be inhomogeneous and possess a "sponge" structure for which the threshold value is higher, and that while supraconductivity is already destroyed by moderate fields in the main part of the structure supraconducting regions still exist with zero induction.

The importance attached to heat capacity measurements below 1° K. has already been mentioned. Simon[59a] has suggested that such measurements might be attempted by combining the material under investigation with the salt employed in the demagnetization method and noting the change this produces in the final temperature. He has also suggested[60] that by using a calorimeter well below 1° K. it should be possible to measure an evolution of heat of the order of 10^{-11} cal./sec. The heat of absorption of X and γ rays could thus be easily measured.

During the past half-dozen years or so a number of low temperature calorimetric investigations have been carried out on diatomic gases by Giauque and his collaborators. Entropy values calculated by means of the third law of thermodynamics have in most cases been found to be in close agreement with values derived from band spectral data. This could be regarded as an experimental verification of the third law. Where there has been lack of agreement it has led to new information concerning the structure of the molecule. Nitrous oxide[61] and ice[62] have now been studied in this way. It is concluded that the linear NNO molecule enters its crystal lattice without much discrimination between its ends, the measured entropy being 1·14 E.U. less than the value obtained from band spectral data, whereas complete lack of discrimination between the ends of the molecule would give a deficiency of 1·377 E.U. Ice is thought to consist of water molecules arranged so that each is surrounded by four others, each molecule being orientated in such a way as to direct its two hydrogen atoms towards two of the four neighbours. The orientations are further restricted by the requirement that only one hydrogen atom lies near each O–O axis.

Ahlberg and Freed[63] have made an experimental determination of the portion of the specific heat associated with electronic transitions in crystals. This has been done by a comparison over the temperature range 17–295° K. of the heat capacities of the two similar rare earth sulphates $Sm_2(SO_4)_3 . 8H_2O$ and $Gd_2(SO_4)_3 . 8H_2O$. There is satisfactory quantitative agreement between the observed and the computed electronic specific heats.

Further progress has been made with the calculation of thermodynamic data from spectroscopic measurements[64, 65, 66, 67, 68] and most of these results have been collected together in a paper by Justi and Lüder[69].

Reverting to purely thermal measurements Saniélévici[70] has used an adiabatic microcalorimeter to measure the heat liberated due to the radioactive radiation from polonium. Zlotowski[71] has measured the heat energies of the β- and γ-radiation

from Ra (B and C), and Swietoslawski and Zlotowski[72] have used a water flow calorimeter to determine the heat generated when γ rays are absorbed.

Determinations of the latent heat of fusion of heavy water have been made by Bartholomé and Clusius[73] using a Bunsen ice calorimeter, and by Jacobs[74], who used a Nernst calorimeter of about $\frac{1}{2}$ gm. capacity. The value found by the former workers for the molecular heat of fusion was 1522 cal. Jacobs used a series of water mixtures and by extrapolation obtained a value of 1525 ± 2 cal. for pure heavy water, and 1524 ± 2 for the 99·7 per cent heavy water used by Bartholomé and Clusius. It is interesting to note that Bernal[75] has suggested a model for the water molecule from which he calculates a value of 1500 cal. for the molar heat of fusion of heavy water.

In order to determine the heats of reaction of the system rubber-sulphur McPherson and Bekkedahl[76] used a calorimeter which consisted of a cylindrical copper block mounted in a 1-litre Dewar flask. This was surrounded by thermal insulation and placed in a double-walled container maintained by means of boiling cymene at a constant temperature of about 175° C.—the temperature of vulcanization. A measured quantity of heat could be added electrically, or removed by inserting steel balls.

A group of American workers[77] have described a gold-plated copper calorimeter, which contains a glass reaction vessel and catalyst in which it is intended to determine the heat changes accompanying catalytic hydrogenations and halogenations. In this way it is hoped to supplement combustion data for substances of high molecular weight where the accuracy of the combustion method decreases. In order to enable the measurements to be made up to 150° C., diethylene glycol is used as calorimetric liquid in the surrounding bath. An automatic arrangement is described for maintaining adiabatic conditions. Results have so far been obtained for ethylene and some of the simpler olefinic hydrocarbons[78]

A small calorimeter has been designed by Woodall[79] to measure directly the Peltier coefficient at the junction of two dissimilar metals. Copper and constantan wires are soldered end to end and enclosed by the calorimeter, which aims at minimizing the amount of heat lost along the leads, and includes a small heating coil wound over the thermojunction.

Jessup[80] has determined the heats of vaporization at 40° C. of eight gasolines used in motor and aircraft engines, and has found that within the accuracy of the experiments the results can be represented as linear functions of either volatility or specific gravity. The method used consisted in vaporizing a known mass of the fuel into a stream of air flowing through a coiled tube immersed in the water of a calorimeter, the rise in temperature of the water being observed.

The heat of combustion of isobutane has been measured by Rossini[81], the same experimental technique being used as for earlier combustion experiments by this author. The value obtained at a temperature of 25° C. is $686{\cdot}31 \pm 0{\cdot}13$ k.-cal./mol., which agrees remarkably well with the value of 685·6 obtained by Thomsen about fifty years ago. Rossini has, in fact, recalculated Thomsen's data by making corrections for the impurities present and obtained a value

of 686·2 ± 1·6, but he considers this agreement to be somewhat fortuitous. The present value for the heat of combustion of isobutane has been used together with the previously reported values for methane, ethane and propane to derive a fairly reliable value for the heat of combustion of neopentane[82]. This value is 840·4 ± 1 k.-cal./mol., which is only a little more than 1 per cent less than the only existing value, also that of Thomsen.

Steinwehr and Schulze have carried out determinations of the heat quantities accompanying transformations in β brass[83], cobalt[84] and iron[85], a similar method being employed in each instance. The specimen in the form of a relatively large block some 20 cm. in height and 10 cm. in diameter, having a thermojunction located at its centre, was electrically heated at its surface and enclosed in a large cement tube packed with heat insulation. The block was cooled from a temperature well above the transformation point and temperature-time curves obtained. Above and below the transformation region straight lines were generally obtained when the logarithm of the temperature was plotted against time. This indicated Newton's law of cooling to hold good. From these curves the increase in the temperature of the block due to the heat evolved as a result of the transformation was obtained, and a measurement of the heat capacity of the apparatus at the mean temperature of the transformation by an electrical input method enabled the quantity of heat due to the transformation to be derived. A similar method using a much smaller specimen enclosed in an evacuated iron crucible has also been used for pure (99·9 per cent) calcium by Schulze[86]. The experimental results are given in table 2.

Table 2

Material	Range of transformation (° C.)	Heat of transformation (cal./g.)
β brass	480 to 430	3·0
Cobalt	1125 to 1070 (magnetic transformation)	1·2 ± 0·07
	420 to 378 (allotropic transformation)	0·1 ± 0·03
Iron	894 to 870 (A_3 point, γ to α transformation)	6·2 ± 0·08
	790 to 710 (A_2 point, magnetic transformation)	4·8 ± 0·2
Calcium	480 to 445	5·3
	307 to 285	—

The results for calcium are interesting in that two years previously the same author in collaboration with Schulte-Overberg had investigated the allotropic transformation of a commercial grade of calcium containing 4·6 per cent of impurities. The method of mixtures was then used, the calcium enclosed in a glass tube filled with argon being dropped into water. Only one transformation was then detected. This occurred at between 450 and 430° C., and the heat associated with it was only 2·3 ± 0·3 cal./gm.

In general, existing data for heats of transformation are somewhat inconsistent. The previous values at the upper cobalt point had been 2·0 and 14·7 cal./gm., whilst those at the A_2 and A_3 iron points had ranged from 3·65 (Umino) to 6·55 (Durrer), and 2·86 (Stansfield) to 6·76 (Oberhoffer and Grosse) respectively.

Differences in composition have been thought to be the cause of some of these variations. Esser and Bungardt[87] however attribute the differences in the results for iron to differences in the actual experimental methods. They show that existing data for the heat content of pure iron become scattered between temperatures of 650 and 900° C., and point out that observers who used methods involving slow rates of cooling obtained high values for the specific heat of iron in this range together with low values for the heat accompanying the γ to α transformation, whereas those who plunged the metal into a water calorimeter obtained low values for the heat content and high values for the heat of transformation. This type of variation is to be expected if rapid cooling prevents complete transformation taking place. Furthermore, Esser, Bungardt and Browne using the water-calorimeter method obtained a value of 6·60 cal./gm. at the A_3 point, whereas in the paper now under consideration reference is made to a slow cool determination made in collaboration with Baerlecken which gave a value of 3·42 cal./gm.

In view of the above suggestion it is interesting to note that Umino[88] has carried out a further extensive series of measurements of the heat content of irons and steels. The water-calorimeter method is used and data are obtained for the heat evolved by some nineteen carbon steels when immersed at gradually increasing temperatures up to above their melting points. Much interesting data are thus obtained, and the fairly high value of 5·59 cal./gm. is given for the heat evolved at the A_3 point. The work of Steinwehr and Schulze previously mentioned does not fit in with Esser and Bungardt's suggestion, since a value of 6·2 cal./gm. was obtained with a relatively slow rate of cooling. It is, of course, possible that the size of specimen may have an effect owing to stresses being set up by differences in temperature.

The specific heats and heats of transformation of iron and steel are quantities of considerable industrial importance; not only do they govern the amount of heat required to effect a steel melt, but also these data, together with those for the density and thermal conductivity, are necessary before calculations of problems concerning the cooling of steel ingots can be attempted. In this connexion Lachmann[89] has made a mathematical analysis of the process of solidification between two parallel planes.

There is great need for further reliable experimental data, and a method of the type recently employed by Sykes[90] in which various rates of heating or cooling can be used might well be adopted for the study of iron, particularly as Naeser[91] claims to have detected no less than six distinct additional discontinuities in the specific heat of iron between 70 and 700° C.

Sykes's method has been devised mainly for the investigation of the type of transformation which consists of a rearrangement of the atoms within the solid solution rather than of an allotropic phase change. Thus in brass having a composition near CuZn, at high temperatures the copper and zinc atoms are arranged in a random manner at the corners and centres of a cube-centred lattice. At a temperature just below 470° C. the copper atoms commence to segregate to the cube corners and the zinc atoms to the centres. There is, however, no sharply

defined temperature at which this process ceases, since with falling temperature the order continues to become more perfect until the temperature becomes too low for further atomic movements.

The amount of energy which should be released due to this ordering of the atoms, and its effect on the cooling curve, has been calculated by Bragg and Williams[92], and others. It is shown that at the temperature at which rearrangement occurs during cooling, the specific heat should increase, and then gradually decrease as the degree of order becomes more perfect. Owing to the influence of the surrounding furnace, the direct cooling methods usually employed cannot well be applied to study such transitions which may extend over 100° C. or more. Sykes[90] suggests two methods for the measurement. The first is a double differential cooling curve method in which comparisons are made of the rates of cooling of the test specimen and a copper standard of the same size when symmetrically enclosed in a copper block. This method suffers from possible errors introduced on account of the difference between the emissivities of the two metals. The second is an absolute method, for the measurement of specific heat as the temperature is continuously varied, which obviates the difficulties due to variations in emissivity. The specimen, A, of mass M, is enclosed inside a copper block, B, which is heated at a slow uniform rate by means of a surrounding furnace. A is fitted with an internal heating coil of small heat capacity, and power input E to this coil is adjusted to bring the temperature of A to within about 0·5° C. below that of B. When adjustments have been made so that the temperatures of both A and B are steadily increasing at the same rate, the input to A is slightly increased to bring the temperature, T_1, of A about 0·5° C. above that of B, T_2. The rate of change of temperature $d(T_1-T_2)/dt$, is obtained when $T_1=T_2$, and from a curve of T_2 against time dT_2/dt is noted at that instant. The specific heat is given by

$$E/M\,[d(T_1-T_2)/dt+dT_2/dt].$$

By varying the conditions so that $d(T_1-T_2)/dt$ has alternate signs the specific heat can be determined at a series of gradually increasing temperatures, and errors due to heat transfer can be eliminated. The main difficulty in the application of this method is to ensure that the zero of the differential thermocouples used to measure the difference between T_1 and T_2 indicates true equality of the two temperatures, since any error from this cause introduces an absolute error in the result which is independent of the sign of $d(T_1-T_2)/dt$.

In trial measurements the specific heat of electrolytic copper was measured over the range 100–860° C. and the experimental values were found to be in close agreement with the determination of Wüst. The specific heat-temperature curve of β brass containing 49·06 per cent zinc was subsequently obtained. It shows a sharp rise from a value of 0·14 at about 470° C. to a maximum of about 0·35 at 462° C., has decreased to 0·16 at 400° C., and gradually continues to decrease to 100° C. It appears that the contribution made to the specific heat by the atomic rearrangement continues over the greater part of this range, probably down to 160° C. There is close agreement between the curves obtained on heating and on cooling.

Bragg and Williams in a subsequent paper[93] have compared these experimental results with their theoretical predictions and also with those of Bethe[94]. The experimental and theoretical specific heat curves are similar in type, but whereas the latter has higher values at low temperatures it does not increase to nearly as high a maximum, that is, the final disappearance of order is more sudden than indicated by the theory. The divergencies balance to some extent, so that energies associated with the order transformation as measured by the areas of the curves above the normal specific heat curve indicated by Dulong and Petit's law are in closer agreement. The calculated energy difference is 11·4 cal./gm. and the observed lies between 10·0 and 8·5 cal./gm. The latter value is obtained if Bethe's estimate of the abnormality in specific heat above the transformation point is used. These values are much in excess of that already mentioned as being obtained by Steinwehr and Schulze. Presumably this is due to the fact that in this latter determination the transformation was considered to be complete within a much narrower range of temperature, and only the increased energy associated with the peak of the curve was measured.

The transformation of Cu_3Au has also been studied, and in this instance there is agreement to within 6 per cent between the total energy change predicted by theory and that found experimentally by Sykes.

Fowler[95] applies somewhat analogous ideas to those of Bragg and Williams to the theory of the rotations of molecules in liquids and solids, and is thereby able to obtain curves approximating to those for the energy content and specific heats of certain halogen hydrides and ammonium salts.

Blackman has contributed a number of theoretical papers dealing with the specific heats of crystals[96, 97, 98, 99]. He has shown that a perfect Debye curve is extremely rare and probably an accidental occurrence. The conditions under which deviations arise are studied, and theoretical justification for the Nernst-Lindemann theory, which represents the vibrational spectrum by two peaks, is established.

References

(48) Laby, T. H. and Hercus, E. O. *Proc. phys. Soc.* **47**, 1003 (1935).
(49) Lang, H. R. and Jessel, R. *J. Instn Petrol. Tech.* **18**, 863 (1932).
(50) Honda, K. and Tokunaga, M. *Sci. Rep. Tôhoku Univ.* **23**, 816 (1934–5).
(51) Poppema, T. J. and Jaeger, F. M. *Proc. K. Akad. Wet. Amst.* **38**, 510 (1935).
(52) Jaeger, F. M., Fonteyne, R. and Rosenbohm, E. *Proc. K. Akad. Wet. Amst.* **38**, 502 (1935).
(53) Keesom, W. H. and Keesom, Miss A. P. *Physica*, 's Grav., **2**, 557 (1935).
(54) Keesom, W. H. and Clark, C. W. *Physica*, 's Grav., **2**, 513 (1935).
(55) Mott, N. F. *Proc. roy. Soc.* **152**, 42 (1935).
(56) Rutgers, A. J. *Physica*, 's Grav., **1**, 1055 (1934).
(57) Mendelssohn, K. and Moore, Miss J. R. *Proc. roy. Soc.* **151**, 334 (1935).
(58) Schubnikow, L. W. and Chotkewitsch, W. J. *Phys. Z. Sowjet.* **6**, 605 (1934).
(59) Gorter, G. J. and Casimir, H. *Phys. Z.* **35**, 963 (1934).
(59*a*) Simon, F. *Not. Proc. roy. Instn*, **28**, 515 (1933–5).
(60) Simon, F. *Nature*, Lond., **135**, 763 (1935).
(61) Blue, R. W. and Giauque, W. F. *J. Amer. chem. Soc.* **57**, 991 (1935).
(62) Pauling, L. *J. Amer. chem. Soc.* **57**, 2680 (1935).

(63) Ahlberg, J. E. and Freed, S. *J. Amer. chem. Soc.* **57**, 431 (1935).
(64) Lewis, B. and von Elbe, G. *J. Amer. chem. Soc.* **57**, 612 (1935). See also p. 2739.
(65) Johnston, H. L. and Walker, Miss M. K. *J. Amer. chem. Soc.* **57**, 682 (1935).
(66) Gordon, A. R. *J. chem. Phys.* **3**, 259 (1935).
(67) Gordon, A. R. *J. chem. Phys.* **3**, 336 (1935).
(68) Vold, R. D. *J. Amer. chem. Soc.* **57**, 1192 (1935).
(69) Justi, E. and Lüder, H. *ForschArb. IngWes.* **6**, 209 (1935).
(70) Saniélévici, A. *J. Phys. Radium*, **6**, 73 (1935).
(71) Zlotowski, I. *J. Phys. Radium*, **6**, 242 (1935).
(72) Swietoslawski, W. and Zlotowski, I. *C.R. Acad. Sci.*, Paris, **200**, 660 (1935).
(73) Bartholomé, E. and Clusius, K. *Z. phys. Chem.* B, **28**, 167 (1935).
(74) Jacobs, L. *Trans. Faraday Soc.* **31**, 813 (1935).
(75) Bernal, J. D. *Nature*, Lond., **135**, 231 (1935).
(76) McPherson, A. T. and Bekkedahl, N. *Bur. Stand. J. Res.*, Wash., **14**, 601 (1935).
(77) Kistiakowsky, G. B., Romeyn, H. Jr., Ruhoff, J. R., Smith, H. A. and Vaughan, W. E. *J. Amer. chem. Soc.* **57**, 65 (1935).
(78) Kistiakowsky, G. B., Ruhoff, J. R., Smith, H. A. and Vaughan, W. E. *J. Amer. chem. Soc.* **57**, 876 (1935).
(79) Woodall, A. J. *Proc. phys. Soc.* **47**, 615 (1935).
(80) Jessup, R. S. *Bur. Stand. J. Res.*, Wash., **15**, 227 (1935).
(81) Rossini, F. D. *Bur. Stand. J. Res.*, Wash., **15**, 357 (1935).
(82) Rossini, F. D. *J. chem. Phys.* **3**, 438 (1935).
(83) von Steinwehr, H. and Schulze, A. *Phys. Z.* **35**, 385 (1935).
(84) von Steinwehr, H. and Schulze, A. *Phys. Z.* **36**, 307 (1935).
(85) von Steinwehr, H. and Schulze, A. *Phys. Z.* **36**, 419 (1935).
(86) Schulze, A. *Phys. Z.* **36**, 595 (1935).
(87) Esser, H. and Bungardt, W. *Arch. Eisenhüttenw.* **8**, 37 (1934–5).
(88) Umino, S. *Sci. Rep. Tôhoku Univ.* **23**, 665 (1935).
(89) Lachmann, K. *Z. angew. Math. Mech.* **15**, 345 (1935).
(90) Sykes, C. *Proc. roy. Soc.* **148**, 422 (1935).
(91) Naeser, G. *Mitt. K.-Wilh.-Inst. Eisenforsch.* **17**, 185 (1935).
(92) Bragg, W. L. and Williams, E. J. *Proc. roy. Soc.* **145**, 699 (1934).
(93) Bragg, W. L. and Williams, E. J. *Proc. roy. Soc.* **151**, 540 (1935).
(94) Bethe, H. A. *Proc. roy. Soc.* **150**, 552 (1935).
(95) Fowler, R. H. *Proc. roy. Soc.* **149**, 1 (1935).
(96) Blackman, M. *Proc. roy. Soc.* **148**, 365 (1935).
(97) Blackman, M. *Proc. roy. Soc.* **148**, 384 (1935).
(98) Blackman, M. *Proc. roy. Soc.* **149**, 117 (1935).
(99) Blackman, M. *Proc. roy. Soc.* **149**, 126 (1935).

§ 3. CONDUCTION

It has long been realized that the measurement of thermal conductivity is one which is particularly prone to experimental errors. Unlike the corresponding electrical case no perfect heat insulator exists and so heat leakage to or from the test specimen is a factor which has invariably to be guarded against. In the case of gases there are additional complications due to convection, radiation and the accommodation effect at a solid surface. These experimental difficulties probably explain why determinations of the conductivity of gases have mostly been limited to measurements in the region of atmospheric temperature.

During last year, however, measurements up to about 300° C. were carried out on hydrogen by Gregory[100], and on carbon dioxide by Archer[101]. The hot-wire

method of the compensated type was used in each instance, the surrounding tube being immersed in a thermostatically controlled oil bath to enable the temperatures to be varied over the required range. Straight lines are obtained when at a constant temperature $(Q/\theta)\ [\log_e (r_2-\lambda')/(r_1+\lambda)]$ is plotted against the reciprocal of the pressure. In the foregoing expression Q is the total heat loss in calories, corrected for radiation; θ the temperature difference between the hot wire and the internal wall; r_1 and r_2 the respective radii of the wire and tube; λ and λ' the mean free paths of the gas molecules at the respective temperatures of the hot wire and the wall of the tube. These straight lines indicate the absence of convection, and the thermal conductivity is derived from the intercept of the line on the axis. The accommodation coefficient can be derived from its inclination.

For carbon dioxide the thermal conductivity is found to be $3{\cdot}47 \times 10^{-5}$ c.g.s. units at 273° K. This is in good agreement with the value of $3{\cdot}43 \times 10^{-5}$ recently obtained by Kannuluik and Martin, and the 1927 value of Weber. The values at higher temperatures are given by $3{\cdot}47 \times 10^{-5}\ (T/273)^{1{\cdot}216}$, T being in degrees absolute.

The conductivity of hydrogen increases from $45{\cdot}9 \times 10^{-5}$ at 295·1° K. to $70{\cdot}8 \times 10^{-5}$ at 593·4° K. From 270 to 502° K. it is represented by $5{\cdot}68 \times 10^{-6}\ T^{0{\cdot}77}$, and above 500° K. by $2{\cdot}2 \times 10^{-4}\ T^{0{\cdot}19}$. The value at 273° K. is about 4 per cent higher than that found by the same author in collaboration with Archer in 1926, but is in closer agreement with the values of Kannuluik and Martin, Weber and Schneider.

In the foregoing method the gas is assumed to obey Maxwell's law, that the thermal conductivity is independent of pressure. In two recent publications attention has been directed to the possibility of exceptions to this law. For such cases the customary method of testing for the absence of convection would become complicated by a true effect of pressure on the conductivity, and it would not be easy to disentangle the two effects.

Ubbelohde[102] has suggested that it is in the case of a polyatomic gas that the thermal conductivity may vary with the pressure, the reason being that at higher pressures any departure from the Maxwell-Boltzmann distribution of energies is more rapidly readjusted. Only when there is full participation of the vibrational energy does the thermal conductivity attain its maximum value, K_{max}, and it is shown that the conductivity K at any pressure is given by $K = K_{max}/(1+F)$, where F depends on the pressure and dimensions of the apparatus. A correlation of the time of relaxation for redistribution of vibrational energy in carbon dioxide as calculated from the dispersion of sound at high frequencies with that calculated from thermal conductivity data is attempted. Reasonable agreement is obtained, although the lack of suitable experimental data precludes an exact quantitative test. The need of further reliable data is apparent, and since it has been found that the presence of small amounts of hydrogen and water vapour can increase the relaxation time by several powers of 10, Ubbelohde suggests that advantage might be taken of this to ascertain whether this has the anticipated effect of increasing K and decreasing the pressure coefficient.

The other contribution in this connexion has been an experimental investigation, by Milverton[103], of thermal conduction through vapours. He used a similar method to that of Gregory and Archer for the study of acetaldehyde vapour and water vapour. A range of pressures was used at several different temperatures for each vapour. The usual test for the absence of convection could not be applied, but the experimental tube was the same as that in which convection had been successfully eliminated for earlier experiments using air; radiation amounted to less than 2 per cent; the accommodation effect could not be measured but the grounds on which its effect is considered small appear to be justified. The experiment showed that Maxwell's law did not hold for either vapour, but should be replaced by a linear conductivity-pressure relation over the ranges studied.

The equation for the thermal conductivity of water vapour in cal./cm.-sec. ° C. is $K=(359{\cdot}8+2{\cdot}170T+3{\cdot}025p-0{\cdot}0270pT)\times 10^{-7}$, where T is the mean temperature of the vapour in ° C., and p is the pressure in cm. of mercury.

The effect of gas pressure will certainly call for careful study in future conductivity work with all but the simple gases.

A method has been used by Eucken and Warrentrup[104] to determine the thermal conductivities of metals when in sheet form. A disc of the metal is electrically heated at its centre and water cooled around the circumference. Independent tests are made to determine the corrections necessary on account of radiation and convection losses from the heater and the specimen itself, and it is finally considered that the thermal conductivity of the light alloy samples, for which the apparatus was designed, can be determined at 0° C. to within about 1 per cent. These workers have realized the importance of making simultaneous measurements of the electrical conductivity, and this quantity is derived from potential difference measurements at points on the disc surface. Values are obtained for the Lorenz function of copper and aluminium which are in close agreement with previously published data. The variation of the thermal and electrical conductivities at 0° C. and − 192° C. of alloys composed of copper, magnesium and aluminium with duration of heat treatments at 100° C. and 215° C. have since been studied[104a]. Both conductivities are found to increase as a result of the treatment but their ratio remains relatively constant.

Reddemann[105] has determined the thermal and electrical conductivities of β-manganese at a temperature of −190° C. The thermal conductivity is of the order of 0·012 c.g.s. unit, and in view of this low conductivity it is not surprising that the Lorenz function has a value of almost three times that indicated by the electronic theory of conduction.

An important contribution to this subject has been made by Smith and Palmer[106], who have determined the two conductivities for over a hundred copper alloys. The longitudinal heat flow method was employed, using rods 0·75 in. in diameter, surrounded by a guard tube, the heat flow being measured by a water flow calorimeter. The experimental points covered temperatures ranging from 40 to 230° C. Values at 20 and 200° C. read from smooth curves drawn through these results are tabulated, and are combined with electrical conductivity measurements at these

two temperatures to give the Lorenz function. The value of the Lorenz function of the purest copper (99·986 per cent) increases from $5{\cdot}45 \times 10^{-9}$ to $5{\cdot}78 \times 10^{-9}$ over this range, and in this type of change it resembles most other pure metals. The Lorenz functions of the alloys have values as high as $8{\cdot}78 \times 10^{-9}$ at 20° C., for alloys containing 20–30 per cent nickel and about 5 per cent zinc (the thermal conductivity being 0·068 c.g.s. unit), but decrease with increase in temperature. This tendency of the Lorenz function to converge towards a more or less common value at high temperatures is similar to that found by the present writer for iron[107] and several heat resistant alloys[108] over the temperature range 0–800° C., and by Lees (1908) and other workers at temperatures below atmospheric. Smith and Palmer by plotting the thermal against the corresponding electrical conductivity, for a given temperature, obtain lines which are approximately straight over about half their length and intersect the thermal conductivity axis at a small but definite positive value. At higher conductivities there is appreciable curvature. When the thermal conductivity is plotted against the product of the corresponding electrical conductivity and absolute temperature all the values for all alloys fall practically on a single curve. This is also straight for a considerable length and bends over at the higher conductivities. Where straight, the curve may be represented by $K = C\lambda T + k$, where K is the total thermal conductivity, λ is the electrical conductivity, T the absolute temperature, k may be regarded as the non-metallic part of the thermal conductivity, and C as a true Lorenz constant for the metallic part of the thermal conductivity. Smith and Palmer find $k = 0{\cdot}018$ and $C = 5{\cdot}71 \times 10^{-9}$ c.g.s. units. It is interesting to note that Schofield's values for pure copper up to 625° C. and Lees for brass, manganin and German silver at temperatures between 18 and −170° C. all fall closely on the curve. The suggestion that this curve can be used to predict the approximate thermal conductivity of a copper alloy from a measurement of its electrical resistance appears to be well founded.

Another group of metals for which a knowledge of their thermal conductivities is of considerable interest and importance is iron and its alloys; and it is gratifying to find that Bollenrath and Bungardt[109] have made an excellent survey of the existing data for these metals. They suggest a value of 0·184 for the thermal conductivity of pure iron at 0° C., whereas Cleaves and Thompson in their recent book[110] suggest a value of approximately 0·19. The thermal conductivity of commercially pure iron was found by the present writer to decrease from a slightly extrapolated value of 0·177 at 0° C. to a value of 0·071 c.g.s. unit at 800° C., and the result was in close agreement with that of Shelton over their common temperature range, 100 to 500° C.

The general nature of the variation of the thermal conductivity of irons and steels can be better appreciated if the data shown in a number of separate diagrams by Bollenrath and Bungardt are all plotted in one figure. It is then seen that whereas at 0° C. the conductivities range from a value of about 0·19 for pure iron to 0·025 for the poorest conducting austenitic steel, at higher temperatures most values, with the exception of Donaldson's, tend to converge so that at 700° C. they only range from 0·05 to 0·08. The preliminary results of Raisch converge to a somewhat

lower mean value of about 0·05. The steels he used were complex tool steels and no details of the method are yet available.

Curves are also given to show the effect of added elements on the thermal conductivity of iron at normal temperatures.

No correlation with electrical conductivities is made, possibly on account of the dearth of data for both conductivities at other than normal temperatures.

Turning to the consideration of the thermal conductivities of refractory materials one should note that several papers on this subject have been read before the American Ceramic Society. Results have been reported for the thermal conductivities of similar silica- and fire-bricks when tested by different workers using different types of apparatus, which are by no means consistent. Those for the silica-brick obtained by Wilkes[111], Austin and Pierce[112] and Finck[113] agree to within 10 per cent for hot-face temperatures of from 200–750° C., but at 1100° C. Wilkes' value is 18 per cent greater than Finck's. The differences are much more marked for the fire-brick, and Finck's values are then the highest. All three used plate types of apparatus, but whereas Finck measured the heat input to the underside of the specimen, the others had their heaters on top and measured the output by a guarded water-flow calorimeter. Heilman and Bradley[114], who used a somewhat crude method in which the test specimen was inserted in the wall of a gas-furnace and faced with a slab of material of assumed conductivity, obtained much lower results. These tests indicate the need for further careful investigation into the methods used.

At the other end of the temperature scale we find that Griffiths and Awbery[115] have studied the thermal conductivities of materials used for solid carbon dioxide containers. The measurements were made on the material in box form, the quantity of heat flowing through the walls being obtained from observation of the rate of evaporation of a block of carbon dioxide inside the box. The following values in c.g.s. units are given for the thermal conductivities at about $-27°$ C., of balsa wood $0{\cdot}00007_5$, expanded rubber $0{\cdot}00006_7$, cork (density 10 lb. per ft^3) $0{\cdot}00008_3$, and cork (density 6·67 lb. per ft^3) $0{\cdot}00007_8$. It is observed that for these materials the thermal conductivity at a fixed temperature is approximately a linear function of the cube root of the density. Furthermore, the value for air falls on the extrapolated line.

The interesting result to which reference was made in last year's report, namely, that the thermal conductivity of silica aerogel is less than that of air, has been followed by a further contribution on the subject by Kistler[116]. The thermal conductivity data previously reported is used to derive an approximate value for the mean free path, L, of the gas molecule within the gel. It is believed that this length is comparable with the size of the voids within the gel, for which there was no previous experimental information. The calculations have been made using conductivity results for three different gases, L being derived from measurements at two different pressures. Table 3 shows the values obtained:

Table 3

Gas	Thermal conductivity (c.g.s. units at 740 mm. pressure and 33·3° C.)	Calculated value of L (cm.)
Air	$4{\cdot}85 \times 10^{-5}$	$4{\cdot}8 \times 10^{-6}$
Carbon dioxide	$4{\cdot}33 \times 10^{-5}$	$5{\cdot}1 \times 10^{-6}$
Dichlorodifluoro-methane	$3{\cdot}71 \times 10^{-5}$	$5{\cdot}1 \times 10^{-6}$

The good agreement for the three different gases shows that to a first approximation at least the average distance apart of the structural elements of this silica aerogel (whose apparent density was 0·184 g./cm^3) must be very close to 5×10^{-6} cm., or 500 A.

References

(100) Gregory, H. S. *Proc. roy. Soc.* **149**, 35 (1935).
(101) Archer, C. T. *Phil. Mag.* **19**, 901 (1935).
(102) Ubbelohde, A. R. *J. chem. Phys.* **3**, 219 (1935).
(103) Milverton, S. W. *Proc. roy. Soc.* **150**, 287 (1935).
(104) Eucken, A. and Warrentrup, H. *Z. techn. Phys.* **16**, 4, 99 (1935).
(104*a*) Eucken, A. and Warrentrup, H. *Z. Elektrochem.* **41**, 331 (1935).
(105) Reddemann, H. *Ann. Phys.*, Lpz., **22**, 1, 28 (1935).
(106) Smith, C. S. and Palmer, E. W. *Tech. Publ. Amer. Inst. Min. Engrs*, No. 648 (1935).
(107) Powell, R. W. *Proc. phys. Soc.* **46**, 659 (1934).
(108) Powell, R. W. *Proc. phys. Soc.* **48**, 381 (1936).
(109) Bollenrath, F. and Bungardt, W. *Arch. Eisenhüttenw.* **9**, 253 (1935–6).
(110) Cleaves, H. E. and Thompson, J. G. *The Metal Iron* (McGraw-Hill Book Co., 1935).
(111) Wilkes, G. B. *J. Amer. ceram. Soc.* **16**, 125 (1933); **17**, 173 (1934).
(112) Austin, J. B. and Pierce, R. H. H. Jr. *J. Amer. ceram. Soc.* **18**, 48 (1935).
(113) Finck, J. L. *J. Amer. ceram. Soc.* **18**, 6 (1935).
(114) Heilman, R. H. and Bradley, R. S. *J. Amer. ceram. Soc.* **18**, 43 (1935).
(115) Griffiths, E. and Awbery, J. H. *Proc. Brit. Ass. Refrigeration*, **31**, 91 (1934–5).
(116) Kistler, S. S. *J. phys. Chem.* **39**, 79 (1935).

§ 4. RADIATION

A knowledge of the rate of gain or loss of heat by the process of radiation is of considerable importance both in industrial problems and in meteorological calculations. It is, therefore, not a little surprising to find that the emissivities of water and ice have only recently been satisfactorily determined. In their book *The Calculation of Heat Transmission* (1932), Fishenden and Saunders give values of 0·63 for ice and 0·67 for water which appear to be based on experiments reported by Siegl in 1907. McAdams, however, in his book on *Heat Transmission* (1933) refers to a personal communication by Hottel and Guerrieri, according to which the emissivity of water increases from 0·95 at 0° C. to 0·96 at 100° C. The present writer considered the emissivity of water to be not far removed from 1 since he found that the emissivity of a linen surface was unaltered when the linen was saturated with water[117]. Recently the matter has received careful investigation at the hands of Schmidt[118], who finds that the emissivity of water in a direction normal to the surface is 0·965. These measurements were carried out on a flowing

stream of water, but it was also found that a film of water only 0·1 mm. thick was sufficient to increase the emissivity of a solid surface from 0·77 to 0·96. Observations on wet ice at 0° C. gave a value of 0·966. Smooth dry ice at −9·6° C. was studied in films 0·4 mm. and 0·8 mm. in thickness when formed on a brass surface. In each case a value of 0·965 was obtained. At the same temperature a film of white hoar frost only 0·1–0·2 mm. thick gave a value of 0·985. Schmidt remarks that such a frost film represents the nearest approach to a dead black surface which has yet been realized. In his experiments the comparative black body was in the form of a hollow cylindrical cavity, the internal walls being painted black and maintained at a known constant temperature by immersing the cylinder in snow or a water bath.

Schmidt also considered the variation of emission coefficient with the angle of emission. As long ago as 1897, Uljanin showed that Lambert's cosine law, according to which the emissivity should be independent of the direction, is not in harmony with Fresnel's reflexion formula. Experiments by Bauer and Moulin on platinum, Worthing on tungsten and carbon, and Henning and Heuse on magnesium oxide have shown the law to be approximately true for angles within 40° of the normal.

For a smooth non-conducting surface, such as ice, the electromagnetic theory of radiation gives $r_n = [(\mu - 1)/(\mu + 1)]^2$, where r_n is the normally reflected fraction, i.e. $r_n = 1 - E_n$, and μ is a mean refractive index for all wave-lengths, equal to $\sin\phi/\sin\psi$, where ϕ is the angle the incident radiation makes with the normal. By means of Fresnel's formula for refraction it follows that

$$r_\phi = 1 - E_\phi = [\sin^2(\phi - \psi)/2\sin^2(\phi + \psi)]\,[1 + \cos^2(\phi + \psi)/\cos^2(\phi - \psi)].$$

From this equation, and the value for μ given by his experiments at normal incidence, Schmidt has calculated E_ϕ for wet ice for all values of ϕ, and found the emission coefficient to remain approximately constant until $\phi = 40°$, to decrease to 0·83 at 70°, and finally, to zero at 90° (grazing incidence). The average emissivity coefficient for all directions is 0·914. Subsequent experiments by Schmidt and Eckert[119] in which the emission was actually measured in the various directions have completely confirmed these predicted values. Similar data are also obtained for wood, glass, paper, clay and copper oxide.

The work was extended to include metallic conductors, and for these it was found that the emissivity definitely increased as the angle with the normal is increased. Thus for polished aluminium at 171° C. the emissivity in the normal direction is 0·0388, at 60° to the normal it is 0·05, and at 87°, 0·18. The average emissivity is 26 per cent greater than that measured in the normal direction.

Advantage has been taken of the low emissivity of aluminium as a means of reducing heat losses in many practical cases. The metal in the form of thin sheets known as "Alfol" is being increasingly incorporated in built-up heat-insulating panels, or applied directly to pipes. It can also be beneficially used as a lining to ground sheets, tent covers and topees.

In another paper Eckert[120] has endeavoured to calculate the manner in which the emissivity of a surface is governed by the actual geometrical form of the surface. Exact solutions of many of the cases considered are difficult to obtain, but curves

are given showing the upper and lower limits for the effective emissivities plotted against the emissivity of the material. The cases include a series of tubes, a battery containing four series of tubes, a surface consisting of right-angled grooves, the space between two parallel infinite semi-planes and cylindrical holes having various ratios of depth to diameter.

On the instrumental side of the subject the claims of the bolometer as a sensitive radiometric device have been revived. Moon[121] has shown that since the bolometer is well adapted to operation on alternating current, its readings can be amplified to give a sensitivity comparable with that obtainable with the best thermocouple and galvanometer. Such an instrument has the advantage that the galvanometer is replaced by a rugged apparatus, and spurious electromotive forces due to slight temperature differences in the circuit have no effect on a tuned detector. Furthermore, when the bolometer strip is included in an a.-c. bridge its resistance has no effect on the detector. This enables long wires of extremely small diameter or sputtered films to be used, and the conduction loss from a small blackened target at the centre of the wire is greatly reduced. A theoretical treatment of such a vacuum target bolometer for use in an a.-c. bridge circuit is given, the assumption being made that the radiation from the surface of the wire is negligible compared with that from the target. For specified dimensions and temperature the sensitivity varies as the square root of the product of the thermal conductivity and the temperature coefficient of resistance of the wire. This quantity is much the same for all pure metals but is considerably smaller for most alloys. Owing to the reductions in radiation losses the sensitivity can be greatly increased by using the bolometer, or any other radiometric instrument, at a lower temperature. Moon's equation shows that at the temperature of solid carbon dioxide the sensitivity is twice that obtainable at room temperature, whilst at the temperature of liquid nitrogen it is about seventeen times as great.

Moon and Mills[122] describe the construction of such a bolometer cell using 40 μ (0·0016 in.) tungsten wire having a target of platinum 2 μ thick and 0·1 × 0·4 cm. in area, the receiving surface being coated with a layer of black evaporated antimony. Maximum sensitivity was not aimed at in this particular instrument, but the results obtained proved the instrument to be satisfactory in operation, and in close agreement with the above-mentioned theory; the latter can now be used as a basis for the construction of a highly sensitive instrument.

References

(117) Powell, R. W. and Griffiths, E. *Trans. Instn chem. Engrs*, Lond., **13**, 185 (1935).
(118) Schmidt, E. *ForschArb. IngWes.* **5**, 1 (1934).
(119) Schmidt, E. and Eckert, E. *ForschArb. IngWes.* **6**, 175 (1935).
(120) Eckert, E. *Arch. Wärmew.* **16**, 135 (1935).
(121) Moon, P. *J. Franklin Inst.* **219**, 17 (1935).
(122) Moon, P. and Mills, W. R. Jr. *Rev. sci. Instrum.* **6**, 8 (1935).

§ 5. CONVECTION AND EVAPORATION

The transference of heat by means of convection is a much more complex process than is that by either conduction or radiation, since it varies according to the properties of the convective fluid and depends on the size and shape of the surface, and also on the character of the fluid flow relative to the surface. The uses and the results of convectional processes are very extensive and a considerable amount of progress has been made in the study of this branch of the subject.

The instability of a layer of fluid when heated from below, which results in the formation of cellular vortices, has received further attention. This was the case Bénard noted early in the twentieth century in the course of experiments using spermaceti. A theory explaining the formation and order of magnitude of these vortices was proposed by Lord Rayleigh in 1916, and Jeffreys later (1928) calculated a critical temperature for the stability of a fluid between two horizontal conducting planes. Schmidt and Milverton[123] have now examined this case experimentally in order to ascertain whether the setting up of the cellular vortices is accompanied by any marked change in the heat transfer coefficient. Brass plates were immersed in water, the lower one being heated by a measured electrical input, and the temperatures of the two plates measured by means of thermocouples. The energy input is plotted against the temperature difference for spacings of 4, 4·5, 5 and 5·5 mm. In each case the curve is at first linear, but a critical temperature difference is later reached at which an increased amount of energy is needed to promote unit rise in temperature. For the distances used these discontinuities occur at temperature differences of 1·7, 1·0, 1·05 and 0·84° C., whereas according to Jeffreys' formula they should occur at about 1·5, 1·1, $0{\cdot}9_7$ and $0{\cdot}8_8$° C. respectively.

The optical method used by Saunders, Fishenden and Mansion[124] for studying convective problems has been applied to this case, photographs being taken both before and after instability has set in. In the stable region the photographs indicated a uniform temperature gradient from the lower to the upper plate, but after the critical temperature difference was passed, although the image still remained perfectly steady, its upper edge had become wavy as would be expected on account of the lateral temperature gradients introduced by the convective cellular motion.

Brunt has demonstrated that for air films the direction of the circulation depends on the film thickness, the flow being downward at the centre for depths of 8 mm. or more, while for shallower cells the flow is upward at the centre.

It has also been suggested that certain cloud formations can be explained as resulting from the instability of a layer of air heated below or cooled above[125, 126]. In the presence of considerable shear the cells are elongated to longitudinal cells parallel to the direction of shear.

In the above-mentioned optical method a parallel beam of light from a point source passes parallel to the heated surface, and after refraction is projected on to a distant screen. Provided the ray remains in the region where the product of the thermal conductivity of the fluid and the temperature gradient therein is constant, quantitative measurements of the heat flow can be derived from its displacement.

In the experiments of Saunders, Fishenden and Mansion on the heat loss from a horizontal plate of approximately 9 × 18 in. area in air and a vertical plate in water the values derived in this way agreed with the measured heat input to within about 6 per cent. An advantage of the method is that it enables the contributions made to the heat loss by separate parts of the surface to be determined. Thus in the air experiments, which ranged up to excess temperatures of the order of 330° C., the heat loss due to natural convection from a heated plate was found to be 60 per cent greater from the upper than from the under surface.

Three Russian workers, Joukovsky, Kirejew and Shamschew[127], have used a similar method but introduced a lens system so as to make the apparatus more compact. They determined the distribution of heat transfer over the curved surface of an electrically heated cylinder when subjected to an air stream at right angles to its axis. Their results show that the heat transfer is a maximum from the element of surface facing the wind, decreases to a minimum at each side, and increases to a maximum at the back. This type of distribution is similar to that found by Lohrisch[128] using a diffusion method; by Paltz and Starr[129] from measurements of the condensation within each of several internal divisions of a pipe supplied with steam; by Klein[130] from observations of the melting of a cylinder of ice; by Small[131] from direct measurements of the heat transfer using a surface thermopile; by Krujilin and Schwab[132] from direct measurements of the external temperature of a steam heated tube, together with an equation for the total coefficient of heat transfer in any particular direction which was obtained by solution of the Laplace equation; and by Powell and Griffiths[133] from measurements of the electrical input to elementary sections of the surface, the latter being maintained at a uniform temperature excess. The last-mentioned experiments were also made with a wet surface, and in this way a similar type of distribution was found for the rate of evaporation around the curved surface of the cylinder.

Reverting to further work on natural convection, in which the only air movement is that induced by the hot body, it is seen that for very fine horizontal wires Borg[134] and Vernotte[135] consider that two distinct laws of convection are obeyed, depending on the temperature excess and wire diameter. This change is probably due to an alternation in the nature of the air movement.

Weise[136] has studied the heat transfer by natural convection from square plates when suspended in horizontal and vertical positions. Two sizes were used, 160 × 160 × 10 mm. and 240 × 240 × 15 mm., and the excess temperatures employed ranged from 50 to 350° C. Over the whole temperature range the heat transfer by convection from both plates is about 20 per cent less in the horizontal than in the vertical position. The results for the coefficient of heat transfer in the vertical position satisfy equations of the type suggested for vertical plane surfaces by Schmidt and Beckmann, $\alpha = 4{\cdot}8\,[(T_w - T_a)/T_a H]^{\frac{1}{4}}\,[B/760]^{\frac{1}{2}}$, and by Nusselt, $\alpha = 2{\cdot}2\,(T_w - T_a)^{\frac{1}{4}}$, except that the numerical coefficients differ. This difference probably results from the finite thickness of the edges of the plates. In the horizontal case thermocouple explorations were made to obtain temperature distributions in the air above and below the smaller plate at plate temperatures of 149 and 246° C.

The total heat transfers deduced from these measurements agree to within about 4 per cent with the measured input. It is concluded that at 149° C. the heat loss from the under side is only 1 or 2 per cent greater than that from the upper side, whilst at 246° C. the difference amounts to about 8 per cent. This result is very different to that previously mentioned as being obtained by Saunders, Fishenden and Mansion. Measurements made at the National Physical Laboratory[137] tend to agree with Saunders's results. For excess temperatures up to 60° C. the heat transfer by convection from the upper surface was about 2·2 and 1·8 times that from the under surface for plates measuring 12 × 12 in. and 36 × 36 in. respectively, when mounted 30 in. above the floor. Naturally the disposition of surrounding bodies can exert a considerable influence. There appears to be room for further work on this aspect of natural convection.

Carpenter and Wassell[138] have studied the convective heat loss as a function of the spacing between two vertical plates and of their height. The plates were each 5 × 4 ft. and some interesting results have been obtained. The total heat loss is decreased when the plates are less than 2 in. apart, and is reduced to about half its initial value when ½ in. apart.

A somewhat similar set of experiments has been carried out at the National Physical Laboratory[139], in which the distance between the plates was varied from 0·1 to 1 cm. and only one of the two plates was heated. The heat loss is again found proportional to a power of the temperature excess above the air, and as the space is reduced the value of this index decreases to unity—the value for conduction. The heat loss was 30 per cent greater with the edges open than when sealed.

Several workers have dealt with the heat transfer to or from fluids flowing in tubes. Among those who obtained results of importance in condenser operation mention may be made of papers by Rhodes and Younger[140] and Jurgensen and Montillon[141]. The former gave data for the heat transfer between various condensing organic vapours and a metal tube, while the latter studied the dependence of heat transfer on the inclination of the tube. Wiesent[142] considers the heat transfer from refrigerating brine solutions to be proportional to the 0·75 power of their viscosities. Schropp[143] has dealt with the effect of dew and hoar frost on the heat transfer to brine pipes from still air. The coefficient was about 60 per cent greater with the dew in drop than in film form. A frost-covered surface gave a value about 10 per cent greater than the corresponding value for a smooth tube.

Kuprianoff[144] and ten Bosch[145] have reviewed existing data and have suggested modifications to the Prandtl equation. Experiments have been conducted at higher flow velocities than previously, Winkler[146] finding that Krassold's equations could be applied up to a Reynolds' number of 500,000; whilst from a theoretical and experimental study of the heat transfer accompanying the rapid flow of gases through tubes a group of Russian workers[147] concluded that up to velocities of approximately 0·95 of that of sound $Nu = 0{\cdot}017 Pe^{0{\cdot}82}$.

Uses are made of convection processes in many branches of science and engineering, notably in heating and ventilation installations. A measurement of

the electrical energy necessary to maintain a fine platinum wire at a fixed temperature is a common means of measuring air velocities. Recent observations by Paeschke[148] and Schubauer[149] show that if accurate results are required allowance must be made for variations in the humidity of the air. The heat loss is greater at higher humidities, and at an air temperature of 25° C. a humidity change from 25 to 70 per cent is equivalent to a change of 6 per cent in the air speed.

Several adaptations of the hot-wire instrument have been made from time to time to enable the direction of the air flow to be measured. A simple device used for this purpose by Powell and Griffiths[150] is that in which a fine butt welded thermocouple mounted parallel to the hot wire, but a small distance downstream, is moved across the air stream until its reading is a maximum. The line joining the wire and thermocouple then gives the required direction. Whytlaw-Gray and Lomax[151] have devised a hot-wire dust precipitator, which is claimed to be very efficient for the particles less than 5 μ in diameter, which are injurious to health. The air under test is streamed slowly through apparatus containing the heated wire and the convection currents, possibly assisted by the Magnus effect, cause the particles to be deposited on a microscope slide. This dust-free zone around a heated wire has also been studied by Miyake[152]. It has often been observed, but does not appear to be fully understood.

The remaining means by which heat transfer can take place from a wet surface is by evaporation, and on account of the large latent heat involved in this process considerable amounts of energy are often transferred in this way. Powell and Griffiths[133], from their study of the rate of evaporation of water from a plane surface 18·2 cm. wide and 24·3 cm. long when subjected to a tangential wind stream, drew attention to the importance of evaporation. They showed that the heat of evaporation amounted to from 70 to 85 per cent of the total transfer at wind speeds of 60–260 cm./sec., for surface temperatures of from 20 to 45° C., air temperature of 20° C., and relative humidities of 40–80 per cent. This paper contains a survey of existing data on evaporation and an experimental investigation of the laws governing the process. The method used was one which enabled the variation in the rate of evaporation over the surface to be determined. The rate was found to be most at the leading edge and to decrease downstream. The total evaporation from the plane surface was proportional to the 0·77 power of the length of surface and increased according to the 0·85 power of the wind speed.

The case of a plane surface subjected to a tangential gas stream was also included by Hoffmann[153] in a series of measurements in which the rate of diffusion from elemental sections of the surface was deduced by a method similar to that previously used by Lohrisch[128]. The surface was covered with strips of filter paper moistened with phosphoric acid, over which ammonia gas was blown at various velocities, and measurements were made of the amount of ammonium phosphate formed. He finds the total diffusion to be proportional to the 0·80 power of both the length and the gas velocity, a result in fair agreement with the foregoing. Hoffmann also studied the diffusion from a small plate inclined at various angles

to the wind stream, from a cylindrical tube having the gas flowing parallel to the axis, either externally or internally, and from various arrangements of batteries of tubes. From the results of these diffusion experiments the coefficient of heat transfer by forced convection is calculated, assuming the truth of the relation between these quantities obtained by Nusselt as a result of dimensional analysis.

Jakob, Erk and Eck[154] have studied the heat transfer resulting from the condensation of water vapour in a vertical tube, and Kirschbaum and Kranz[155] have investigated the complementary case of a vertical evaporating tube. Jakob and Linke[156] have measured the heat transfer from an horizontal plate immersed in a liquid, the only heat loss being by evaporation from the surface of the liquid. In these experiments information has been obtained showing how the size and rate of formation of bubbles affect the heat transfer.

For the purpose of determining the manner in which the rate of evaporation from animals and insects depends on their environment Ramsay[157] has conducted a series of wind tunnel experiments on a model consisting of a small porous tube. He also simulated the tracheal system of an insect by using an open-mouthed cylinder, and found that the variation with wind speed in rate of evaporation from a cylinder of depth equal to about five times the diameter was similar to that actually found in experiments on a cockroach[158]. An interesting discovery was that this insect appears to cover its body with a very thin film of fatty substance which reduces the rate of evaporation at normal temperatures. At a temperature of about 30° C. this substance suffers a change of state and the rate of evaporation assumes a much higher value.

Blüttner[159] has considered the general problem of heat transfer by conduction, convection, evaporation and radiation, particularly in connexion with heat losses from human beings and from the surface of the earth. Experiments are described in which the convectional and evaporational losses are derived from measurements of the temperature and humidity gradients within the laminar film close to the surface.

References

(123) Schmidt, R. J. and Milverton, S. W. *Proc. roy. Soc.* **152**, 586 (1935).
(124) Saunders, O. A., Fishenden, M. and Mansion, H. D. *Engineering*, **139**, 483 (1935).
(125) Phillips, G. C. and Walker, G. T. *Quart. J. R. met. Soc.* **58**, 23 (1932).
(126) Walker, G. T. *Quart. J. R. met. Soc.* **59**, 389 (1933).
(127) Joukovsky, V., Kirejew, A. and Shamschew, L. *Techn. Phys. U.S.S.R.* **1**, 4, 353 (1935).
(128) Lohrisch, W. *Forschungsarbeiten*, No. 322 (1929).
(129) Paltz and Starr. Thesis in Chem. Eng. Mass. Inst. Techn. (Referred to by Drew and Ryan, Amer. Instn Chem. Engrs, June 1931.)
(130) Klein, V. *Arch. Wärmew.* (June 1934).
(131) Small, J. *Phil. Mag.* **19**, 251 (1935).
(132) Krujilin, G. and Schwab, B. *Techn. Phys. U.S.S.R.* **2**, 4, 312 (1935).
(133) Powell, R. W. and Griffiths, E. *Trans. Instn chem. Engrs*, **13**, 175 (1935).
(134) Borg, C. *C.R. Acad. Sci.*, Paris, **200**, 1918 (1935).
(135) Vernotte, P. *C.R. Acad. Sci.*, Paris, **200**, 1919 (1935).
(136) Weise, R. *ForschArb. IngWes.* **6**, 281 (1935).

(137) *Rep. nat. phys. Lab.*, Lond., p. 33 (1935). (H.M. Stationery Office.)
(138) CARPENTER, L. G. and WASSELL, H. C. *Proc. Instn mech. Engrs*, **128**, 439 (1934).
(139) *Rep. nat. phys. Lab.*, Lond., p. 33 (1935).
(140) RHODES, F. H. and YOUNGER, K. R. *Industr. Engng Chem.* **27**, 957 (1935).
(141) JURGENSEN, F. D. Jr. and MONTILLON, G. H. *Industr. Engng Chem.* **27**, 1466 (1935).
(142) WIESENT, J. *Z. Kälte Ind.* **42**, 187 (1935).
(143) SCHROPP, K. *Z. Kälte Ind.* **43**, 81 (1935).
(144) KUPRIANOFF, J. *Z. techn. Phys.* **16**, 1, 13 (1935).
(145) TEN BOSCH, M. *Z. techn. Phys.* **16**, 4, 105 (1935).
(146) WINKLER, K. *ForschArb. IngWes.* **6**, 261 (1935).
(147) GUCHMANN, A., ILJUCHIN, N., TARASSOWA, W. and WARSCHAWSKI, G. *Techn. Phys. U.S.S.R.* **2**, 5, 375 (1935).
(148) PAESCHKE, W. *Phys. Z.* **36**, 564 (1935).
(149) SCHUBAUER, G. B. *Bur. Stand. J. Res.*, Wash., **15**, 573 (1935).
(150) POWELL, R. W. and GRIFFITHS, E. *J. Heating Ventilating Engrs*, **3**, No. 25, 20 (1935).
(151) GREEN, H. L. and WATSON, H. H. *Physical Methods for the Estimation of Dust Hazard in Industry*. H.M. Stationery Office (1935).
(152) MIYAKE, S. *Rept. Tôhoku Univ. Aero. Res. Inst.* No. 123 (June 1935).
(153) HOFFMANN, W. *ForschArb. IngWes.* **6**, 293 (1935).
(154) JAKOB, M., ERK, S. and ECK, H. *Phys. Z.* **36**, 73 (1935).
(155) KIRSCHBAUM, E. and KRANZ, B. *ForschArb. IngWes.* **6** Forschungsheft, No. 375 (1935).
(156) JAKOB, M. and LINKE, W. *Phys. Z.* **36**, 267 (1935).
(157) RAMSAY, J. A. *J. exp. Biol.* **12**, 355 (1935).
(158) RAMSAY, J. A. *J. exp. Biol.* **12**, 373 (1935).
(159) BLÜTTNER, K. *Veröff. pruss. met. Inst.* **10**, No. 5 (1934).

§ 6. HYGROMETRY

A book[160] has recently been published which reviews the question of the control of humidity under industrial conditions and deals with the principles and mode of operation of the equipment used for this purpose.

Lieneweg[161] has considered the problem of the direct determination of the water-vapour content of gases. The dew point depends directly on this quantity, and it is shown that at temperatures above 30° C., the curves representing constant dew point, obtained when the psychrometric temperature difference is plotted against the air temperature, are a series of parallel straight lines. It thus becomes a simple matter to deduce the dew point from wet- and dry-bulb temperatures, and by incorporating wet- and dry-resistance thermometers in a simple electrical bridge circuit an instrument is obtained which can be operated from a distance to give a direct reading of the absolute humidity. At temperatures below 30° C. the connexion between the dew point and the psychrometric temperature difference becomes a little more complicated, but it is still possible to devise the appropriate electrical circuit to provide a direct-reading instrument suitable for observations at air temperatures of from 5 to 30° C.

The hair hygrometer is treated in an analogous way. The curves of equal dew point, obtained by plotting the extension of a hair (which is a function of the relative humidity) against the air temperature, approximate at the higher humidities to parallel straight lines over the range −30 to 40° C. By linking this instrument with an appropriate electrical circuit, direct readings of the vapour content can be

obtained over a wide range of temperatures provided the humidity is not too low.

In recent years the problem of the measurement of humidity at temperatures below 0° C. has become one of considerable importance on account of the study which is being made of the atmospheric condition of cold storage rooms and refrigerated holds. Owing to the decreasing amount of vapour present, and the uncertainty of the saturation moisture content, these measurements become increasingly difficult at low temperatures. Of the available methods the most suitable for routine measurements are the wet- and dry-bulb thermometer in one of its forms and hygrometers of the extending hair type. The performance of the latter at low temperatures is not yet fully understood, but appears to depend on the previous history of the instrument. An investigation of the wet- and dry-bulb hygrometer has recently been carried out by Awbery and Griffiths[162] over the temperature range -2 to $-19°$ C. The temperatures were measured by means of mercury thermometers, resistance thermometers and thermocouples. The actual humidity was generally obtained from dew-point measurements, but from time to time gravimetric determinations were made. As a result of the measurements a new relative humidity table has been prepared. There is good agreement with existing tables except for depressions of less than 1° C. at temperatures below about $-9°$ C. Within this region the new table indicates a definitely lower value for the relative humidity than is given by the *Aspirationen Psychrometer Tafeln*, published by the Prussian Meteorological Institute.

The dew-point apparatus used in the foregoing work consisted of a silver mirror cooled by circulation of pre-cooled paraffin. A thermocouple was used to measure the temperature of the silver surface.

A set of psychrometric charts has been published by Brooks[163], which will be of assistance in the interpretation of data obtained at pressures differing from normal. They cover a range of pressures of from 0·125 to 10 atmospheres, and are based on a formula of the well-known Ferrel type, but the constants employed have been adjusted to conform with recent data.

References

(160) Marsh, M. C. *Controlled Humidity in Industry* (Chas. Griffin and Co., Ltd., London, 1935).
(161) Lieneweg, F. *Wiss. Veröff. Siemens-Werken*, **14**, 20 (1935).
(162) Awbery, J. H. and Griffiths, E. *Proc. phys. Soc.* **47**, 684 (1935).
(163) Brooks, D. B. *Misc. Publ. U.S. Bur. Stand.* No. 146 (Jan. 1935).

§ 7. DILATATION

One very important application of the dilatational properties of materials in connexion with the subject of heat has been their utilization as a means of measuring temperatures. A matter of somewhat wider interest is that of temperature control, and Himmler[164] has published descriptions of a number of temperature regulators which operate by the expansion of solids, liquids or gases.

The practical use which is made of expansion in such processes as shrinking on of metal tyres is well known, but it is interesting to note that cooling of the centre piece can at times be just as effective. Thus in the construction of a bascule bridge in New Jersey, four bearing hubs each weighing 6390 lb. were contracted by being totally immersed in a mixture of solid carbon dioxide and alcohol, before insertion in the girders[165]. On withdrawal from the solution the alcohol film lasted long enough to prevent the formation of frost.

In another large engineering scheme—the Boulder Dam—cracking due to expansion caused by the liberation of heat of chemical action during the setting of cement has had to be avoided. This has been done by installing a refrigerating plant and embedding pipes in the centre of the concrete through which cold water or brine could be pumped. In order to economize in the power necessary for this cooling, preliminary experiments were carried out, as a result of which it was found that by reducing the aluminate and increasing the silicate in the cement the production of chemical heat, particularly in the early stages, was reduced by about 30 per cent[166].

Nearly forty years have elapsed since Guillaume discovered invar, and until quite recently no alloy having a smaller coefficient of expansion at atmospheric temperatures has been available. Masumoto[167] has lately been studying ternary alloys composed of cobalt, iron and chromium, and has found that alloys containing 36–36·5 per cent iron, 54–54·5 per cent cobalt and 9–10 per cent chromium have smaller coefficients than invar. It is, in fact, possible to obtain a zero coefficient or even one having a small negative value. As with invar, the coefficient only assumes a minimum value in the neighbourhood of atmospheric temperature, and may reach about 12×10^{-6} at 200° C. Owing to the fact that the new alloy is practically incorrodible in dilute sodium chloride solution it has been named "stainless invar".

Silver iodide is another substance which contracts on heating, although the values previously given for its coefficient ranged from $-2{\cdot}7 \times 10^{-6}$ to $-55{\cdot}7 \times 10^{-6}$. A smaller value of $-1{\cdot}6 \times 10^{-6}$ has now been obtained by a method in which a new form of differential weight thermometer was employed[168].

Austin and Pierce, who a few years ago developed a vacuum dilatometer for use up to high temperatures[169], in which the change in length was measured by means of an interferometer, have now[170] studied a series of iron-chromium alloys containing from 3·25 to 10·58 per cent of the latter element. As with pure iron the expansion coefficient has a maximum value at about 450° C., but no minimum is observed at the Curie point.

Hidnert[171] has measured the expansion coefficients of a number of single crystals of antimony, and three samples of polycrystalline antimony at various temperatures between 20 and 560° C. It speaks well for the work carried out by Fizeau, as long ago as 1869, that the values now found for the coefficient of expansion between 20 and 100° C. in directions along and perpendicular to the trigonal axis are $17{\cdot}2 \times 10^{-6}$ and $8{\cdot}0 \times 10^{-6}$, whereas he found $16{\cdot}8 \times 10^{-6}$ and $9{\cdot}0 \times 10^{-6}$.

Expansion coefficients have also been reported for sodium tungstate[172], soda-

alumina-silica glasses[173], single crystals of mercury[174], cast and rolled zinc[175], and copper from 75 to 1356° K.[176] Bridgman[177] has extended his measurements of the compressibility and thermal expansion of lithium, sodium and potassium up to 20,000 kg./cm.^{2} The interesting result is obtained that whereas at normal pressures potassium has the greatest, and lithium the smallest expansion coefficient, these relative positions are reversed at high pressures.

References

(164) Himmler, C. R. *Arch. tech. Messen*, **5**, T 151 (1935).
(165) Brown, A. W. and David, E. V. *Steel*, **96**, No. 2 (1935).
(166) Robertson, G. R. *Industr. Engng Chem.* **27**, 242 (1935).
(167) Masumoto, H. *Sci. Rep. Tôhoku Univ.* **23**, 265 (1934–5).
(168) Jones, G. and Jelen, F. C. *J. Amer. chem. Soc.* **57**, 2532 (1935).
(169) Austin, J. B. and Pierce, R. H. H. Jr. *Physics*, **3**, 240 (1932).
(170) Austin, J. B. and Pierce, R. H. H. Jr. *Trans. Amer. Inst. min.* (*metall.*) *Engrs*, Iron and Steel Division, **116**, 289 (1935).
(171) Hidnert, P. *Bur. Stand. J. Res.*, Wash., **14**, 523 (1935).
(172) Austin, J. B. and Pierce, R. H. H. Jr. *J. chem. Phys.* **3**, 683 (1935).
(173) Faick, C. A., Young, J. C., Hubbard, D. and Finn, A. N. *Bur. Stand. J. Res.*, Wash., **14**, 133 (1935).
(174) Hill, D. M. *Phys. Rev.* **48**, 620 (1935).
(175) Jones, H. G. *Proc. phys. Soc.* **47**, 1117 (1935).
(176) Haring, W. J. and Davey, W. P. *Phys. Rev.* **47**, 337 (1935).
(177) Bridgman, P. W. *Proc. Amer. Acad. Arts Sci.* **70**, 71 (1935).

MAGNETISM

By L. F. BATES, D.Sc., Ph.D., F.Inst.P.,
Professor of Physics, University College, Nottingham

§ 1. INTRODUCTION

In the preceding volumes of *Reports on Progress in Physics* some attention has been paid to the theoretical aspects of modern magnetism, and with the appearance of the articles by Sommerfeld and Bethe [1] and the publication of *Magnetism and Matter* [2] many of the theoretical contributions of recent years have been conveniently summarized. There remains, however, a very large body of work of fundamental importance, mainly experimental, which will be surveyed as fully as possible in the present report. An indication of the amount of this work may be gained by consulting the list of some 211 references given in a review of research work in magnetism during 1933–34 compiled by Spooner [3], although work on dia- and paramagnetics has been generally excluded from this list. In the present report attention will be in the main confined to work carried out in the period 1930–35, but, because this is an initial report and a report on magnetism will not be published annually, some latitude will be taken in the inclusion of material published outside this interval of time.

The writer has obviously been forced to make a somewhat arbitrary selection of material, and some branches of the subject have perforce been omitted. For example, magneto-optics and galvanomagnetic phenomena are not discussed here, but these and other effects are well treated elsewhere.

References

(1) Sommerfeld, A. and Bethe, H. *Geiger-Scheel Handbuch der Physik*, **24**, part 2, 333 (1933).
(2) Stoner, E. C. *Magnetism and Matter* (Methuen, 1935).
(3) Spooner, T. *Elect. Engng*, **54**, 1354 (1935).

§ 2. THE PRODUCTION AND MEASUREMENT OF MAGNETIC FIELDS

Details of several new magnets have recently been made available. For example, Häder [4] has described the construction of a very large magnet made for use at Leyden. Cockcroft [5] has designed a special annular magnet, with a gap 80 cm. in mean diameter, 5 cm. in radial width and 1 cm. across, in which uniform fields of the order of 18,000 oersteds can be produced, for the examination of the α-rays emitted by radio-active substances. Lawrence and Livingston [6] have used a magnet with pole-faces 27·5 in. in diameter in a recent form of their apparatus for the production of high-speed ions by multiple acceleration in a magnetic field. The total weight of iron used in the magnet circuit is about 65 tons and the weight of the exciting coils about 9 tons.

Permanent magnets have recently found important applications. Cockcroft, Ellis and Kershaw[7] have designed a permanent magnet for α-ray spectroscopy. From a rigid U-shaped yoke of soft iron are supported two horizontal arms each consisting of 23 laminated plates of cobalt steel, bolted together and fitted with pole-pieces of high permeability steel. The latter have plane faces 500 sq. cm. in area and give a maximum field of 2300 oersteds when placed 5·5 cm. apart. Field coils are placed on the horizontal arms, and the field can easily be changed and set at new permanent values by the passage of suitable exciting currents of short duration. Briggs has described the construction of a permanent magnet suitable for α- and β-ray spectroscopy which gives a field of 5300 oersteds between flat pole-pieces $26 \times 17{\cdot}4$ cm., and 1·19 cm. apart, and Surugue[8] has discussed the construction of large permanent magnets for β-ray work. The use of permanent magnets is strongly to be advised when fields have to be kept constant over long periods, but steel containing 35–36 per cent cobalt is expensive. Special alloys for permanent magnets are discussed by Honda, Masumoto and Shirakawa; their latest material, containing 15–35 per cent cobalt, 10–25 per cent nickel, 8–25 per cent titanium and the remainder iron, may exhibit a coercivity of 800–900 oersteds with a residual induction of about 7500 gauss. It is specially suitable for short magnets but is very difficult to forge. Such materials must possess at the same time high coercivity H_c and remanence B_r, for it is easy to show that the volume of material required in the magnet is a minimum when $H^1.B$ is a maximum, H^1 being the field acting in the material; hence the product $H_c.B_r$ must also be as large as possible. Jellinghaus[8] reports that an alloy of 50 per cent cobalt and platinum possesses a very high value of $H_c.B_r$, equal to 12×10^6 units, H_c being 4000 oersteds.

Special solenoids. Iron-free solenoids specially designed to give large uniform fields have been described by Campbell, Houston[9], Kinsler[10], and by Gerloff and Löwe[11]. The former used a solenoid 80 cm. long, consisting of 18 layers of no. 4 square-section cotton-covered wire. The layers were separated by fibre spacers, and cooled oil circulated through them. In this way a current of 200 amperes at 270 V. produced a steady field of 7300 oersteds over a length of 6 cm. occupied by a source of light whose Zeeman effect was under investigation.

In a study of cosmic-ray tracks Kunze[12] produced a field of 18,000 oersteds over the region occupied by a Wilson chamber by means of a coil supplied with a current of 1000 amperes at 500 V. taken at night from the Rostock town station. The field, however, could only be maintained for about 50 sec., after which the coil required a day in which to cool. The stray field in this case is considerable and loose pieces of iron are sent hurtling through the air when the current is switched on.

Field-control and stabilization. In many experiments the field of an electromagnet can be maintained sufficiently steady if the current is kept constant. Potter[13] has described a simple accumulator circuit which enables this to be done. The magnetizing current supplied from the d.-c. mains or from a main laboratory battery is stabilized by taking a small compensating current from a bank of accumulators of small capacity. The main current passes through an adjustable resistance R, a fixed resistance S and the magnet coils. The stabilizing current

passes through a portion of S and the magnet coils, and, if this portion is chosen so that practically no current is taken from the bank of accumulators, then about 90 per cent of the fluctuation of the mains supply is neutralized.

Rutherford, Wynn-Williams, Lewis and Bowden[14] have used a fluxmeter with a mirror attached to the moving coil to control and measure changes in the magnetic field of the annular magnet mentioned above. Two large coils are mounted in the gap and joined in series with the fluxmeter and a magnetic compensator. The latter consists of two Helmholtz coils between which is mounted a permanent magnet, whose position may be changed, causing a change in flux which will bring the fluxmeter coil to any desired position. The compensator can be calibrated in terms of the gap field and thus used for the measurement of small changes. The method of control just described does not permit one to deal with slow continuous changes of magnetic field, such as may be produced by a gradual change in temperature of the magnet core. Wynn-Williams[15] has therefore used the device in conjunction with an automatic field stabilizer. A narrow, rectangular spot of light is reflected from the fluxmeter mirror and falls upon a double-cathode photo-electric cell so that both cathodes are partly covered when the spot is in its zero position. Each cathode is connected through a simple valve and thyratron circuit, entirely worked from a.-c. mains, to a system of relays. A movement of the spot of light causes the one system of relays or the other to be operated and to set in motion a series motor. The latter slides a movable contact in a mercury resistance until the electromagnet current is altered by the amount necessary to restore the fluxmeter coil to its equilibrium position. Arrangements are made to deal with large changes of field, to prevent over-correction and to provide protection against failure of the current supplies or of the lamp.

Breakdown rectifiers have been used by Dupouy[16] to protect field coils against the dangers of short circuit or failure of the field current supply. He has also designed an instrument termed a gaussmeter, for the measurement of magnetic fields; it depends for its action upon the relatively large couples which may act upon magnetically anisotropic crystals such as siderose ($FeCO_3$) and oligist (Fe_2O_3) when placed in a uniform field. These couples are directly proportional to the field strength, and Dupouy has constructed a robust form of instrument, but it would appear that strict temperature control is essential. An accurate and simple method of measuring strong, extensive magnetic fields to a few parts in 10^5 by means of an ordinary physical balance is described by Briggs and Harper[17].

References

(4) Häder, G. *Siemens-Z.* **10**, 481 (1930).
(5) Cockcroft, J. D. *J. sci. Instrum.* **10**, 71 (1933).
(6) Lawrence, E. O. and Livingston, M. S. *Phys. Rev.* **45**, 608 (1934).
(7) Cockcroft, J. D., Ellis, C. D. and Kershaw, H. *Proc. roy. Soc.* A, **135**, 628 (1932).
(8) Surugue, J. *J. Phys. Radium*, **6**, 94 (1935).
Jellinghaus, W. *Z. tech. Phys.* **17**, 33 (1936).
(9) Campbell, J. S. and Houston, W. V. *Phys. Rev.* **39**, 601 (1932).
(10) Kinsler, L. E. and Houston, W. V. *Phys. Rev.* **45**, 104 (1934).

(11) GERLOFF, G. and LÖWE, E. *Z. Phys.* **98**, 559 (1936).
(12) KUNZE, P. *Z. Phys.* **79**, 203 (1932) and **80**, 559 (1933).
(13) POTTER, H. H. *J. sci. Instrum.* **11**, 95 (1934).
(14) RUTHERFORD, Lord, WYNN-WILLIAMS, C. E., LEWIS, W. B. and BOWDEN, B. V. *Proc. roy. Soc.* A, **139**, 617 (1933).
(15) WYNN-WILLIAMS, C. E. *Proc. roy. Soc.* A, **145**, 250 (1934).
(16) DUPOUY, M. G. *Ann. Phys.*, Paris, **14**, 549 (1931).
(17) BRIGGS, G. H. and HARPER, A. F. A. *J. sci. Instrum.* **13**, 119 (1936).

§ 3. SUSCEPTIBILITY MEASUREMENTS

Standard Determinations. In view of the many comparison measurements of magnetic susceptibility in which water is used as a standard substance, an accurate absolute determination of its susceptibility by Auer[18] is important. He used a modified form of the Quincke method which requires that one column of the liquid under investigation shall occupy a vertical tube of uniform cross-section, with the meniscus between the flat pole-pieces of an electromagnet, the tube being joined to a reservoir remote from the magnet. On exciting the uniform field the meniscus rises or falls and is usually restored to its initial position by adjustment of the level in the reservoir by suitable means. In Auer's apparatus the liquid is contained in a rigid bent tube system which is firmly clamped to a rectangular brass frame. The latter can be rotated about a horizontal axis until the meniscus in the limb remote from the field occupies a standard position in the field of view of a microscope fixed to the frame. From the rotation necessary to restore the meniscus to its standard position when the field is excited, the change in level h and therefore the magnetic force on the column in the known magnetic field H can be found. The air above the liquid is replaced by a diamagnetic gas and the essential portions of the tube system are water-jacketed, in order that changes in density may not arise. In Auer's arrangement the density of the liquid to be used in the calculation of the magnetic susceptibility is the density of the column within the field, and not that of the column outside, as is the case in many earlier determinations. The mass susceptibility is given by the expression

$$\chi = \frac{2.g.h}{H^2 - H_0^2}\left\{1 - \frac{\rho_1}{\rho}\right\} + \frac{k_1}{\rho},$$

where H_0 is the stray field acting on the lower end of the column in the field, ρ the density of the liquid, and ρ_1 and k_1 respectively the density and volume susceptibilities of the gas above the liquid. Auer's value for the susceptibility of water is $-0{\cdot}72145 \pm 0{\cdot}00048 \times 10^{-6}$ e.m.u. per gram, with a temperature coefficient between 15 and 20° C. equal to $1{\cdot}45 \times 10^{-4}$ per degree.

Incidentally, Cabrera and Fahlenbrach[39], and Spencer and Trew[40] find that heavy water, deuterium oxide, has a mass susceptibility of $-0{\cdot}637 \pm 0{\cdot}001 \times 10^{-6}$, which gives a molar susceptibility of $-12{\cdot}76 \times 10^{-6}$ on the assumption that the molecular weight of D_2O is 20·027. Taking Auer's value for water, Vogt[19] has recently found the mass susceptibility of mercury to be $-0{\cdot}168 \times 10^{-6}$ at 20° C., a value which has been confirmed by Tai and the writer[19] in the course of some

experiments on mercury amalgams. This new value is important, for mercury can be purified without great difficulty, and gives us a standard liquid with about three times the volume susceptibility of water. The value generally accepted for mercury up to the present has been $-0{\cdot}190\times10^{-6}$.

Vogt has also given a very full discussion of the correction to be applied to eliminate the effects of ferromagnetic impurity when measurements are made by the Gouy method. Usually this is done by the Owen-Honda method, using the equation $(k)_H=(k)_\infty+I_s/H$, where $(k)_H$ is the observed value of the volume susceptibility when the field between the pole-pieces is H, $(k)_\infty$ the theoretical value in an infinitely large field and I_s is the saturation magnetization of the ferromagnetic impurity. The correct form of the last equation to be applied to Gouy measurements only is

$$(k)_H=(k)_\infty+\frac{2I_s}{H+H_0}-\frac{2D}{H^2-H_0^2},$$

where D is the area between the I, H curve for the impurity and the ordinate at $H=H_0$, the value of the stray field at the end of the body remote from the pole-pieces. Fortunately, it is rarely necessary to use the term containing D, but the introduction of the quantity $2I_s$ in place of I_s is important when an estimate of the amount of impurity present is to be made.

A new method for the measurement of magnetic susceptibilities has been devised by Kapitza and Webster[20] for use with test materials which are exposed to mechanical stresses; this method has been incorporated by Potter in some important measurements described on page 197. Rankine[21] has described a new method in which the susceptibilities are measured in very weak fields; it requires for its proper performance an iron-free laboratory. Kapitza[22] has also designed a special form of magnetic balance for use with very intense magnetic fields of short duration.

Measurements on gases. The measurement of the magnetic susceptibility of a diamagnetic gas is usually somewhat difficult because of the small susceptibility of the uncompressed gas and because of the serious effects which a trace of the strongly paramagnetic oxygen may produce. Hence, comparative measurements are always made in these cases, and the experimental methods are mainly elaborations of the method originally designed by Faraday for comparing the susceptibilities of different substances in the form of rods. Faraday suspended each rod in turn from the same torsion fibre so that it lay between the poles of an electromagnet, and he noted the magnitude and direction of the rotation which occurred when the field was excited. In modern experiments on the susceptibility of gases the rod is replaced by a test-body of special design, which, when the field is excited, is brought to its standard position by turning a torsion head to which the fibre is attached. If the angles through which the torsion head is turned when the test-body is in vacuum and in a gas are respectively θ_0 and θ_1, then $(\theta_1-\theta_0)$ is a measure of the susceptibility of the gas.

Special test-bodies have been devised by Bitter[23], Hammar[24] and others, a peculiarly simple form being described by Rao and Sivaramakrishnan[25]. The

most extensive study of test-bodies has been made by Havens[26] and the form which he considers most satisfactory is shown in figure 1.

It consists of two evacuated spherical bulbs of quartz connected together, and compensated by a small quartz rod. The whole system weighs only 0·1 g., and is coated with gold so that it may be earthed to avoid electrostatic effects. Special attention is paid to the centring of the torsion fibre with respect to the torsion head, and very high sensitivity is attained; e.g. with oxygen at a pressure of 18·3 cm. of mercury the twist of the fibre was 454·65 complete turns for a field of 7300 oersteds; it is presumed that the elastic limit was not exceeded. Some of his results are given in table 1; they are undoubtedly the most accurate so far obtained.

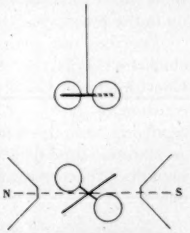


Figure 1. Test-body for measurements of susceptibilities of gases. (Havens.)

The magnetic properties of nitric oxide, NO, are of considerable interest, for Van Vleck[27] has shown that at absolute zero the isolated gas molecule, which is then in a $^2\Pi_{\frac{1}{2}}$ state, should possess zero magnetic moment, whereas at high temperatures, where it is in the $^2\Pi_{\frac{3}{2}}$ state, it should possess a moment of nearly 2 Bohr magnetons. Experimental

Table 1. Molar susceptibilities χ_M of gases (Havens) $\times 10^6$

H_2	He	Ne	A	CO_2	N_2	O_2
−4·0051	−1·906	−7·651	−19·23	−20·88	−11·938	+3335

determinations in the temperature range 135–300° K. have been made by Aharoni and Scherrer[28], Bitter[29], Stössel[30] and Wiersma[31], and excellent agreement between theory and experiment has been obtained. Lips[32] has recently made measurements on solid NO, where he finds but a weak paramagnetism of about $0{\cdot}0198 \times 10^{-3}$ per g.-mol., which is constant between 63 and 90° K. within the limits of experimental error, and is attributed to the formation of double molecules whose resultant spin is zero. The last result is not, of course, in agreement with the theoretical values calculated from the Van Vleck theory, and, as the lattice structure of NO is unknown the susceptibility for the solid state cannot be calculated.

Few experiments have been made with diamagnetic vapours, either because of their violent action on containers or the lack of knowledge of the relevant vapour densities. Gerlach[33] has published some preliminary results of measurements of the susceptibility of potassium vapour obtained with a modified form of Lehrer's[34] balanced column method. The Lehrer method has been used by Lues[35], who finds that oxygen does not obey the Curie law exactly between 300 and 600° K., and by Mann[36], who measured the susceptibility of oxygen in fields ranging from 1800 to 15,000 oersteds and found it constant. Mann's paper gives many details of the Lehrer method. Mann also used this method with rare gases and found that a linear relation existed between ionization potential and susceptibility in these cases.

Experiments on paramagnetic vapours have been made by Lallemand[37] and Néel[38] who enclosed a weighed quantity of the parent substance inside a small evacuated bulb with stout walls, and measured the magnetic forces upon the bulb and its contents when placed in a non-uniform field by means of a translation balance with electromagnetic control. Néel used sulphur vapour, and Lallemand used anhydrous iron perchloride, $FeCl_3$. The molar susceptibility of the latter decreased from $9{\cdot}98 \times 10^{-3}$ at 125° to $4{\cdot}59 \times 10^{-3}$ at 327° C. where it vaporized, and the susceptibility thereafter remained constant up to 430° C., after which it again decreased with rise in temperature. The difficulties of weighing the small quantities of the material required in these experiments is apparent, and, in any case, the chemical constitution of the vapour is sometimes a matter of surmise.

References

(18) Auer, H. *Ann. Phys.*, Lpz., **18**, 593 (1933).
(19) Vogt, E. *Ann. Phys.*, Lpz., **14**, 1 (1932).
Bates, L. F. and Tai, L. C. *Proc. phys. Soc.* **48**, 795 (1936).
(20) Kapitza, P. and Webster, W. L. *Proc. roy. Soc.* A, **132**, 441 (1931).
(21) Rankine, A. O. *Proc. phys. Soc.* **46**, 391 (1934).
(22) Kapitza, P. *Proc. roy. Soc.* A, **131**, 224 (1931); see also Auwers, O. v. *Naturwissenschaften*, **24**, 65 (1936).
(23) Bitter, F. *Phys. Rev.* **35**, 1572 (1930).
(24) Hammar, G. W. *Proc. nat. Acad. Sci.*, Wash., **12**, 597 (1926).
(25) Rao, S. R. and Sivaramakrishnan, G. *Proc. phys. Soc.* **46**, 318 (1934).
(26) Havens, G. G. *Phys. Rev.* **41**, 337 (1932) and **43**, 992 (1933).
(27) Van Vleck, J. H. *The Theory of Electric and Magnetic Susceptibilities* (Oxford, 1932).
(28) Aharoni, J. P. and Scherrer, P. *Z. Phys.* **58**, 749 (1929).
(29) Bitter, F. *Proc. nat. Acad. Sci.*, Wash., **15**, 638 (1929).
(30) Stössel, R. *Ann. Phys.*, Lpz., **10**, 393 (1931).
(31) Wiersma, de Haas and Capel. *Commun. phys. Lab. Univ. Leiden*, No. 212 b.
(32) Lips, E. *Helv. phys. Acta*, **8**, 247 (1935).
(33) Gerlach, W. *Rep. Int. Congr. Phys. Como* (1927).
(34) Lehrer, E. *Ann. Phys.*, Lpz., **81**, 1129 (1927).
(35) Lues, H. *Z. Phys.* **98**, 537 (1936).
(36) Mann, K. E. *Z. Phys.* **98**, 548 (1936).
(37) Lallemand, A. *C.R. Acad. Sci.*, Paris, **194**, 1726 (1932)
(38) Néel, L. *C.R. Acad. Sci.*, Paris, **194**, 2035 (1932).
(39) Cabrera, B. and Fahlenbrach, H. *Naturwissenschaften*, **22**, 417 (1934).
(40) Spencer, J. F. and Trew, V. C. G. *Nature*, Lond., **137**, 497 (1936).

§ 4. MAGNECRYSTALLINE PROPERTIES

We now turn to experiments on single crystals, in which the direction of magnetization does not, in general, coincide with the direction of the applied field, although in any crystal there exist three mutually perpendicular, principal magnetic axes along which an applied field produces magnetization in the same direction. The susceptibilities along these three axes are termed the principal susceptibilities and are denoted by k_1, k_2 and k_3 or χ_1, χ_2 and χ_3 in volume and mass measurements respectively. Measurements with powdered material provide a value of $\frac{1}{3}(\chi_1+\chi_2+\chi_3)$. When the directions of the magnetic axes are known we may apply

a field with a gradient parallel to each in turn and measure each principal susceptibility directly, provided that the shape of the crystal does not produce spurious volume effects. The latter can be avoided by cutting the crystal to a spherical shape. This direct method was used by Jackson[41] in an investigation of the principal susceptibilities of paramagnetic crystals down to liquid-air temperatures. The crystal was attached to a quartz frame carrier with a standard face of the crystal vertical and perpendicular to the horizontal field of an electromagnet. The frame was suspended from a Sucksmith ring balance which measured the vertical forces, $k_1 . H . dH/dx$, etc.

Often, however, as in the case of a monoclinic crystal, the direction of only one principal magnetic axis, k_3, is known. A circular disc of the crystal may then be cut with the plane of the disc perpendicular to the known axis, and, when suspended with this plane horizontal and parallel to a uniform field H, the disc will be acted upon by a couple $\frac{1}{2}(k_1 - k_2) . \sin 2\phi . H^2 . v$, where ϕ is the angle between the direction of the field and the k_1 axis and v the volume of the disc. For $\phi = 45°$ the couple is a maximum, so that by finding the position in which the maximum couple is exerted, the directions k_1 and k_2 are fixed and their difference can be measured. The two methods described above can, of course, be combined, when the measurements give the ratio k_1/k_2 and the difference $(k_1 - k_2)$. This procedure was adopted by Dupouy[42]. Sometimes, however, couples cannot readily be measured and only translational forces can be utilized. The determination of k_1 and k_2 then becomes much more complicated, since measurements of the forces in a given direction for at least three different positions of the disc with respect to the field are required, as in Jackson's[43] earlier work.

The effects of any surrounding medium have been omitted from the preceding discussion, but, clearly, when the surrounding medium has a volume susceptibility equal to k_1, say, then no translational force can act on the crystal parallel to the k_1 axis. Hence, if the crystal be immersed in a solution whose concentration can be adjusted until the force parallel to the k_1 axis vanishes, k_1 must then be equal to the known susceptibility of the solution. This method was used by Rabi[44], and it possesses the advantage that the crystal need not be very accurately shaped.

Finally, a crystal may be suspended by a torsion fibre parallel to the k_3 axis, when it may oscillate with a period T_0. If, now, the period of oscillation becomes T_1 when a uniform field H is applied parallel to the k_1 axis, then

$$k_1 - k_2 = \frac{T_0^2 - T_1^2}{T_1^2} \frac{c}{H^2} \frac{1}{v},$$

or

$$\chi_1 - \chi_2 = \frac{T_0^2 - T_1^2}{T_1^2} \frac{c}{H^2} \frac{1}{m},$$

where c is the torsion constant of the fibre and m the mass of the crystal. This method was first used by Stenger[45] and by König[46], and has been extensively employed by Krishnan and his collaborators[47]. The latter attach the crystal to the lower end of a very short piece of glass fibre, and suspend it by a vertical quartz fibre from a torsion head, one magnetic axis of the crystal being parallel to the axis

of suspension. They assume that no complications arise from the shape of the crystal, and special attention is paid to the uniformity of the applied field. The absolute value of the greatest susceptibility is found by the solution method already described. The oscillation method is not satisfactory when the magnetic anisotropy is so small that effects of asymmetry of shape cannot be neglected. Krishnan[48] then immerses the crystal in a liquid whose volume susceptibility is equal to that of the powdered crystal, and measures the couples acting upon it in a uniform field. This procedure was very satisfactory, for example, in the case of $MnSO_4.(NH_4)_2SO_4.6H_2O$, where $\chi_1-\chi_2=11{\cdot}4\times10^{-6}$, $\chi_1-\chi_3=6{\cdot}8\times10^{-6}$, while $\frac{1}{3}(\chi_1+\chi_2+\chi_3)$ is $13{,}830\times10^{-6}$ e.m.u. per g.-mol. Krishnan and Banerjee[49] have further described methods for dealing with minute crystals.

Among the more important results obtained may be noted Jackson's results with crystals containing ions in which only electron spins effectively contribute to the magnetic moment; these were found magnetically isotropic to within about 1 per cent. Cobalt sulphate ($CoSO_4.7H_2O$) is peculiar in that the curves of $1/\chi_1$ etc. against temperature are linear but not parallel, and potassium ferricyanide shows a rather complicated behaviour. Dupouy found that the principal susceptibilities of siderose ($FeCO_3$) and dialogite ($MnCO_3$) obeyed Curie-Weiss laws, but oligist (Fe_2O_3) was peculiar in that χ_1 and χ_2 both increased with temperature to a maximum in the neighbourhood of 60° C. and then decreased with further rise in temperature. For a discussion of the relation between the magnetic properties and the X-ray structure of crystals reference should be made to the works quoted above.

Owing to the rapid development of the technique for the production of metallic single crystals, long rods each formed of a single crystal can now be obtained in the case of many metals. If we suppose that a crystal axis is parallel to the long axis of the rod, then the latter may be suspended vertically from one arm of a sensitive balance as in the Gouy method for measuring susceptibility. The magnetic force acting on the rod when one end is in a field H is equal to

$$(k_1\cos^2\theta+k_2\sin^2\theta)\frac{\alpha}{2}H^2$$

or

$$\{\chi_2+(\chi_1-\chi_2)\cos^2\theta\}\frac{\alpha}{2}\rho H^2,$$

where α is the area of cross-section of the rod, ρ its density, and θ is the angle which the k_1 axis makes with the line of centres of the pole-pieces. Clearly, this force depends on θ, and the flexible connexion commonly used to suspend the rod must be replaced by a rigid connexion in which a small aluminium turn-table or torsion head is inserted, so that the specimen may conveniently be orientated. This arrangement has been used by Focke[50], Goetz and Focke[51] and Vogt[52]. Focke showed that, when the principal crystalline axis of a bismuth crystal was parallel to the long axis of the rod, χ_1 was equal to χ_2; in fact, the mass susceptibility parallel to the principal axis, $\chi_\parallel$, was $-1{\cdot}053\times10^{-6}$ while that in a plane perpendicular thereto, $\chi_\perp$, was $-1{\cdot}482\times10^{-6}$. The magnetic anisotropy $\chi_\perp/\chi_\parallel$ was therefore 1·425.

Since we have in effect only two principal axes to deal with, it is not essential that a crystalline axis coincide with the long axis of the rod. For, when the principal crystalline axis makes an angle ϕ with the long axis the expression for the magnetic force becomes

$$\{(\chi_{\parallel} \sin^2 \phi + \chi_{\perp} \cos^2 \phi) \cos^2 \theta + \chi_{\perp} \sin^2 \theta\} \frac{\alpha}{2} \rho . H^2,$$

so that peak values are obtained at $\theta = 0$ and $\theta = 90$, giving the values of

$$(\chi_{\parallel} \sin^2 \phi + \chi_{\perp} \cos^2 \phi)$$

and $\chi_{\perp}$ respectively, whence the two susceptibilities can be calculated if ϕ is known. In general, however, it is best to use several crystals with different values of ϕ in order to avoid directional uncertainties concerning $\chi_{\parallel}$ and $\chi_{\perp}$.

On account of the pronounced diamagnetism of bismuth, Goetz and Focke[51] have made an intensive study of the effects of added impurity on these properties. Added impurity is effective only when soluble in bismuth, when the effects are very marked. The magnetic anisotropy is increased when the impurity is electropositive with respect to bismuth and decreased when it is electronegative. The absolute values of the susceptibilities are reduced in all directions and large changes are found in the temperature variation of the anisotropy. Thus, on plotting $\chi_{\perp}/\chi_{\parallel}$ as a function of temperature in the case of a bismuth crystal containing 0·005 per cent of lead, two hyperbolic curves are obtained, the ratio becoming infinite in the neighbourhood of 0° C. and taking on negative values at lower temperatures, when the crystal is paramagnetic in one direction and diamagnetic in the other. When the reciprocal $\chi_{\parallel}/\chi_{\perp}$ of the anisotropy is plotted as a function of the temperature for a series of crystals containing known amounts of a given impurity a series of straight lines is obtained. These, when produced, intersect in one point at an absolute temperature T_0 above the melting point of bismuth, and each line obeys a relation of the form

$$\chi_{\parallel}/\chi_{\perp} = R = R_0 - (\gamma_0 + \gamma)(T_0 - T),$$

where R_0 and γ_0 refer to pure bismuth, T is the absolute temperature, and γ is positive for metals electropositive to bismuth and negative for electronegative metals. For further relations and details of other interesting phenomena and their interpretation on Ehrenfest's large electron orbit hypothesis, the original paper should be consulted. Another, and probably more adequate, explanation of some of the phenomena on the basis of the formation of Brillouin zones in the alloys has been put forward by Jones[53]. Hoge[54] has extended the work to tin single crystals containing small quantities of antimony, cadmium and gallium, using a sensitive microbalance because of the small volume susceptibility of white tin. Table 2 gives the values of the principal atomic susceptibilities so far obtained with single metal crystals.

Single crystals of ferromagnetics, of course, form an interesting study on their own, but nearly the whole of the most important work on them was carried out prior to the period covered by this report, and little can therefore be summarized here; a very good account is given by McKeehan[55]. Shih[56] has studied the properties of single crystals of iron-cobalt using a pendulum magnetometer designed by

Table 2. Principal atomic susceptibilities of single crystals

Metal	$-(\chi_{\parallel})_A \times 10^6$	$-(\chi_{\perp})_A \times 10^6$	Difference $(\chi_{\perp})_A - (\chi_{\parallel})_A \times 10^6$	Magnetic anisotropy $\chi_{\perp}/\chi_{\parallel}$
Zn	12·4	9·5	+ 2·9	0·766
Cd	29·4	18·0	+11·4	0·612
Hg	22·5	24·3	− 1·8	1·080
Bi	219·0	308·0	+89·0	1·405
Sn	20·9	23·4	− 2·5	1·120

McKeehan[57]. In these body-centred cubic lattices the direction of easiest magnetization was found to change from the [100] to the [111] direction as the percentage of cobalt increased. This indicates that the direction of easiest magnetization does not depend on the crystal structure alone, but also upon the type of atom of which it is constituted. The remanence I_r exhibited by single crystals of iron has been shown by Kaya[58] to obey the relation $I_r = I_\infty/(l+m+n)$, where l, m and n are the direction cosines of the external field with respect to the edges of the crystal cube; this relation appears to warrant fuller examination.

References

(41) Jackson, L. C. *Proc. roy. Soc.* A, **140**, 695 (1933).
(42) Dupouy, M. G. *Ann. Phys.*, Paris, **15**, 495 (1931).
(43) Jackson, L. C. *Philos. Trans.* A, **224**, 1 (1923).
(44) Rabi, I. I. *Phys. Rev.* **29**, 174 (1927).
(45) Stenger, F. *Ann. Phys.*, Lpz., **20**, 304 (1883) and **35**, 331 (1888).
(46) König, W. *Ann. Phys.*, Lpz., **31**, 273 (1887).
(47) Krishnan, K. S., Guha, B. C. and Banerjee, S. *Philos. Trans.* A, **231**, 235 (1933).
(48) Krishnan, K. S., Chakravorty, N. C. and Banerjee, S. *Philos. Trans.* A, **232**, 99 (1933).
(49) Krishnan, K. S. and Banerjee, S. *Philos. Trans.* A, **234**, 265 (1935).
(50) Focke, A. B. *Phys. Rev.* **36**, 319 (1930).
(51) Goetz, A. and Focke, A. B. *Phys. Rev.* **45**, 170 (1934).
(52) Vogt, E. *Ann. Phys.*, Lpz., **21**, 791 (1934).
(53) Jones, H. *Proc. roy. Soc.* A, **147**, 396 (1934).
(54) Hoge, H. J. *Phys. Rev.* **48**, 615 (1935).
(55) McKeehan, L. W. "Ferromagnetism in Metallic Crystals." *Tech. Publ. Amer. Inst. Min. Engrs*, No. 554 (1934).
(56) Shih, J. W. *Phys. Rev.* **46**, 139 (1935).
(57) McKeehan, L. W. *Rev. sci. Instrum.* **5**, 265 (1934).
(58) Kaya, S. *Z. Phys.* **84**, 705 (1933).

§ 5. EXPERIMENTS WITH SINGLE PARTICLES

Few experiments of the Stern-Gerlach type have been carried out within the last five years, but an important precision determination of the numerical value of the Bohr magneton by Meissner and Scheffers[59] must be noted. The success of their work is due to the efficient manner in which the conditions prescribed by Leu[60] for the production of very narrow but intense beams of atomic rays were

applied, and to the method, first devised by Taylor[61], employed in their detection. The apparatus is shown diagrammatically in figure 2.

It consisted of two brass compartments C and R. The first contained the oven O in which potassium or lithium was evaporated, and it was joined to R by the brass tube T, T. The whole apparatus was very efficiently evacuated, and water-cooled blocks W, W, condensed stray alkali atoms. The evaporated atoms passed through the narrow oven slit, through the fine collimating slit S and on to the receiver N, being exposed to the action of a non-uniform field of the Stern-Gerlach type during their passage through T, T. Inside N was a stretched tungsten filament 0·02 mm. in diameter, so mounted that it could be moved across the incident beams of atomic rays, parallel to the slits. It was heated electrically, and alkali atoms falling upon it lost electrons and could leave the surface as positive ions, the resulting ionization current between the filament and a collecting electrode or plate being a measure of the number of atoms received in unit time.

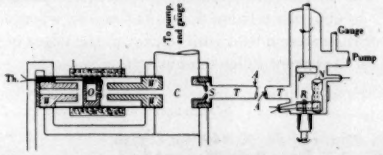


Figure 2. Apparatus for the measurement of the Bohr magneton. (Meissner and Scheffers.)

Special precautions were taken to avoid fluctuations of the oven temperature, and the gradient of the non-uniform field was carefully mapped by means of a tiny search coil mounted on a brass tube, fitted to a cathetometer with horizontal and vertical movements. The initial beam of alkali atoms was split into two by the action of the field, the distributions of the atoms in the deflected and undeflected beams being evaluated from the ionization-current measurements, while the separation of two beams into which an infinitely thin initial beam of single velocity rays would be split was computed on the basis of the theoretical treatment given by Stern[62].

The final values obtained for the Bohr magneton from 4 sets of determinations with potassium and 5 sets with lithium were $0{\cdot}916 \pm 0{\cdot}005$ and $0{\cdot}916 \pm 0{\cdot}003 \times 10^{-20}$ respectively, the mean being $0{\cdot}916 \times 10^{-20}$ erg gauss, correct to 5 parts in 1000.

Experiments with atomic hydrogen were made in 1927 by Wrede[63] and by Phipps and Taylor[64], who showed that the free hydrogen atom possessed a magnetic moment of 1 Bohr magneton. Special difficulties are encountered when experiments are made with diatomic gases, which may be appreciated by considering the behaviour of oxygen. The oxygen molecule normally exists in a $^3\Sigma$ state, i.e. the orbital electron momentum Λ along the line of centres of the constituent atoms is zero, while the electron spin momentum S is $h/2\pi$, so that the total momentum J is $1h/2\pi$-unit. Hence, on the basis of its magnetic properties alone an oxygen molecule would be expected to set in $2J+1$ discrete positions when placed in a

magnetic field, thus giving one undisplaced and two displaced beams in a Stern-Gerlach experiment. Unfortunately, the molecule rotates rapidly about the line of centres and there is a very strong coupling between the spin and the molecular rotation. A reasonable value to take for the most probable rotational momentum l is $9h/2\pi$-units, which would give $\Sigma(2S+1)(2l+1)$ or 165 beams in a Stern-Gerlach experiment. In other words, so many discrete positions are possible that the classical conditions of an infinite number of orientations in a field would be simulated; consequently, the Stern-Gerlach record would be a broad band and very difficult to interpret. Such a band was obtained in the case of a beam of sulphur vapour molecules by Shaw and Phipps[65], and was considered to provide proof that the sulphur molecule possessed a magnetic moment of 2 Bohr magnetons, corresponding to a $^3\Sigma$ state.

A comprehensive series of experiments on O_2 and NO molecular beams have been made by Schnurmann[66] using an arrangement very similar to that designed by Knauer and Stern[67] for measurements with light gases, and shown in figure 3.

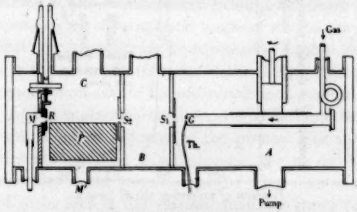


Figure 3. Apparatus for molecular-ray measurements with light gases. (Schnurmann.)

The gas entered an evacuated chamber A through the fine slit G, its temperature being measured by a thermocouple, and some molecules passed through the fine slits S_1 and S_2 to form the molecular beam which passed through a non-uniform field of the Rabi[68] type between pole-pieces P. The broadened beam then fell upon a receiving slit R and part of it entered the tube M. R was parallel to the other slits, and could be moved across the beam. M communicated with a chamber containing a Pirani hot-wire manometer, and the molecules passing through R caused the pressure to rise therein until the number of molecules entering was equal to the number leaving; thus the Pirani gauge measured the relative numbers of molecules in the various portions of the beam, while a compensating gauge attached to M' eliminated fluctuations due to changes in the pressure of gas in the chamber C. The pattern or distribution of molecules in the beam varied greatly with the Rabi field and the temperature of the beam. Curve I in figure 4 shows the pattern obtained with zero field and Curve II the pattern obtained with an initial beam of oxygen molecules of section BB passing, as shown, above the pole-pieces NS, the temperature being 300° K. The strong asymmetry of Curve II is due to the non-uniformity of the field, for while the undeflected molecules moved in this case

in a field of 6800 oersteds, those corresponding to the left-hand portion moved in a stronger field than those corresponding to the right-hand portion, so that the degrees of coupling of electronic and molecular spins were different. Following a procedure due to Einaudi[69], the pattern to be expected on the assumption of fixed values of S and l with the type of Rabi field actually used was calculated and compared with the experimental pattern. The agreement between experiment and theory, with the above values of J and l, is good.

The rotational effects are even more complicated in the case of the NO molecule. The normal state of the molecule is $^2\Pi_{\frac{1}{2}}$ where the spin and orbital momenta are antiparallel and the molecule is non-magnetic. At high temperatures many molecules pass into the $^2\Pi_{\frac{3}{2}}$ state which exhibits a magnetic moment of 2 Bohr magnetons.

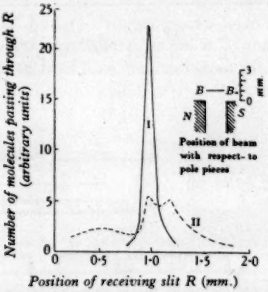


Figure 4. Schnurmann's results for oxygen beams. Curve I, pattern for no field. Curve II, pattern for mean field of 6800 oersteds in region of undeviated beam of section *BB*.

Molecules in this state should give rise to 5 beams in a Stern-Gerlach or Rabi experiment if rotational effects are absent. Actually, on account of these effects it is advisable to use strong fields in which decoupling occurs and $\Sigma 2\,(2J+1)$ magnetically deflected and one undeflected beam of non-magnetic molecules should be obtained, the most probable value of J the total momentum about the line of centres being $\frac{1}{2}h/2\pi$-units. So many components could not possibly be unravelled, but a relatively simple procedure is possible both for O_2 and NO. As before, the molecule is assumed to have a specified magnetic moment and known partition of the rotational states. The shape of the beam pattern curve and its maximum intensity I_0 are obtained from the Pirani gauge readings in the absence of a magnetic field, and the intensity of the undisplaced component I is calculated for known conditions of the magnetic field and compared with that obtained by experiment. For example, in the case of oxygen the experimental value for I/I_0 was 0·287, whilst calculation based on $l=1, \ldots 15$ gave 0·299. On the whole, agreement between experiment and theory is as good as can be expected.

The magnetic methods of measuring nuclear spins and nuclear magnetic moments have already been summarized by Feather[70] and will not be further

discussed here. It may, however, be pointed out that measurements of the magnetic moment of the proton were made by Fisch and Stern[71] and by Estermann and Stern[72], with an apparatus very much like that shown in figure 3, with a Stern-Gerlach instead of a Rabi type of field for producing the magnetic deflections. These direct measurements are possible with hydrogen, because the molecule is diamagnetic and therefore possesses no large permanent magnetic moment. Unfortunately, ordinary hydrogen is a mixture of parahydrogen in which the two protons are set with their spins antiparallel and orthohydrogen in which they are set with their spins parallel with an associated magnetic moment of $2\mu_n$, or 2 nuclear magnetons (1 nuclear magneton = 1/1840 Bohr magneton). At very low temperatures, about 99 per cent of the parahydrogen molecules possess no rotational energy, but the orthohydrogen molecule possesses at least $1h/2\pi$-unit of momentum even at liquid-air temperatures, although the coupling between the rotation and nuclear spins is very small and is readily destroyed in a Stern-Gerlach field. The pattern for a beam of orthohydrogen molecules should therefore, theoretically, consist of two deflected and one undeflected beam, each beam consisting of three components corresponding to the magnetic moment due to the molecular rotation being aligned parallel or perpendicular or antiparallel to that due to nuclear spin. The determination of the magnetic moment of the proton thus requires the determination of the magnetic moments of the orthohydrogen molecule and the magnetic moment $l.\mu_R$ due to the molecular rotation, the difference between them being twice the value of the magnetic moment of the proton. The quantity μ_R is determined from experiments on pure parahydrogen, for at 90° K. this consists of 98·3 per cent of molecules in the state $l=0$ and 1·7 per cent in the state $l=2$. The relative numbers of molecules in the various quantum states can be calculated at higher temperatures from Boltzmann's formula. The maximum intensity of the undisplaced pattern when the field is applied can be compared with the value calculated from an assumed value of μ_R and the distribution in the several quantum states at the temperature of the experiment and a satisfactory estimate thus made. Fortunately, the value of μ_R is relatively unimportant. A similar experiment is then performed with ordinary hydrogen and the maximum intensity of the undisplaced beam when the field is applied is compared with that calculated on the assumptions that the relative percentages of ortho- and parahydrogen molecules are known, that the orthohydrogen molecule possesses a magnetic moment $n'.\mu_n$, where n' is an integer, and that the distribution of rotational states in the beam is known. In this manner μ_R was found equal to $0{\cdot}9\mu_n$, a value agreeing with that calculated by Wick[73], and the nuclear magnetic moment of the orthohydrogen molecule equal to $5\mu_n$, whence that of the proton was $2{\cdot}5\mu_n$. The experimental error is given as about 10 per cent, and it is clear that if theoretical arguments are to be based on this value, its extremely complex experimental basis must be remembered. Moreover, the value obtained by Rabi, Kellogg and Zacharias[74] is $3{\cdot}25 \pm 0{\cdot}3\mu_n$. Estermann and Stern[75] also attempted to measure the nuclear magnetic moment of the deuteron, but owing to experimental and other difficulties beyond their control their results can only be taken as indicating that the nuclear

magnetic moment of the deuteron is considerably less than that of the proton. Rabi and his collaborators have also found the nuclear magnetic moment of the deuteron to be $0{\cdot}75 \pm 0{\cdot}2\,\mu_n$, but, as they point out, they could not settle the sign of the moment by these experiments; they have just announced[76], however, that they have devised a method, involving the use of non-adiabatic transitions in a weak rotating magnetic field, which shows that the moments of both proton and deuteron are positive. The small value of the magnetic moment of the proton has been discussed by Born[77] on the basis of the postulate, due to Pryce, that the positron, proton and electron are different electromagnetic states of the same elementary charge.

References

(59) Meissner, W. and Scheffers, H. *Phys. Z.* **34**, 48 and 245 (1933).
(60) Leu, A. *Z. Phys.* **41**, 551 (1927).
(61) Taylor, J. B. *Z. Phys.* **57**, 242 (1929).
(62) Stern, O. *Z. Phys.* **41**, 563 (1927).
(63) Wrede, E. *Z. Phys.* **41**, 569 (1927).
(64) Phipps, T. E. and Taylor, J. B. *Phys. Rev.* **29**, 309 (1927).
(65) Shaw, E. J. and Phipps, T. E. *Phys. Rev.* **38**, 174 (1931).
(66) Schnurmann, R. *Z. Phys.* **85**, 213 (1933) and *J. Phys. Radium*, **6**, 99 (1935).
(67) Knauer, F. and Stern, O. *Z. Phys.* **53**, 766 (1929).
(68) Rabi, I. I. *Z. Phys.* **54**, 190 (1929).
(69) Einaudi, R. *R.C. Accad. Lincei*, **16**, 132 (1932).
(70) Feather, N. *Reports on Progress in Physics*, **2**, 81 (1936).
(71) Fisch, R. and Stern, O. *Z. Phys.* **85**, 4 (1933).
(72) Estermann, I. and Stern, O. *Z. Phys.* **85**, 17 (1933).
(73) Wick, G. C. *Z. Phys.* **85**, 25 (1933).
(74) Rabi, I. I., Kellogg, J. M. B. and Zacharias, J. R. *Phys. Rev.* **46**, 157 (1934).
(75) Estermann, I. and Stern, O. *Z. Phys.* **86**, 132 (1933).
(76) Rabi, I. I., Kellogg, J. M. B. and Zacharias, J. R. *Nature*, Lond., **137**, 658 (1936). Rabi, I. I. *Phys. Rev.* **49**, 324 (1936).
(77) Born, M. *Mem. Ind. Inst. Sci.* **2**, 554 (1935).

§ 6. GYROMAGNETIC PHENOMENA

As Auwers[78] has recently shown, it is possible to divide gyromagnetic experiments into eight different groups, but in practice it is necessary only to consider two of them. One consists in the rotation of a cylinder of any substance about its long axis when mounted in a field-free space, when, on account of the Larmor precession of the electrons, the cylinder becomes magnetized in a direction parallel to the axis of rotation. Such magnetization can, of course, be measured only in the case of ferromagnetic materials, and the experiments made by Barnett since 1914 have recently been summarized[79]. The other group of experiments consists in the measurement of the change in angular momentum which accompanies a change in magnetization. If we take the elementary magnetic particle to be an electron system then its angular momentum U is equal to $j.h/2\pi$, and the corresponding magnetic moment μ is equal to $g.j.h/2\pi$, so that

$$\frac{U}{\mu} = \rho_R = \frac{1}{g}\frac{2m}{e},$$

where g is the Landé splitting factor for the system. In the case of a simple orbital motion $g=1$, while for a spinning electron $g=2$. The experimental values of $\rho_R \frac{e}{m}$ for iron and iron oxides are given in table 3.

Table 3. Gyromagnetic ratio for iron and iron oxides

Substance	$\rho_R.e/m$	Method	Authors
Iron	1·00 ± 0·01	Direct	Chattock and Bates[80]
,,	1·00 ± 0·01	Null	Sucksmith and Bates[81]
,,	1·00 ± 0·01 ?	Controlled resonance	Coeterier[82]
Iron oxides	1·00 ± 0·01	Simple resonance	Ray-Chaudhuri[83]
Iron	1·05 ± 0·02	Magnetization by rotation	Barnett[79]
,,	1·031 ± 0·003	Null	Barnett[79]

The experimental values of $\rho_R.e/m$ for other substances are given in table 4.

Table 4. Gyromagnetic ratios

Substance	$\rho_R.e/m$	Method	Authors
Cobalt	1·03 ± 0·06	Null	Sucksmith[84]
	1·096 ± 0·002	,,	Barnett[79]
Nickel	1·01 ± 0·01	Direct	Chattock and Bates[80]
	1·00 ± 0·01	Null	Sucksmith and Bates[81]
	1·05 ± 0·01	,,	Barnett[79]
Heusler alloy	1·00 ± 0·01	,,	Sucksmith and Bates[81]
	1·00 ± 0·01	,,	Barnett[79]
Pyrrhotite	3·17 ± 0·03	Controlled resonance	Coeterier[82]

From these tables it is clear that the value of g for all ferromagnetics except pyrrhotite, and possibly cobalt, is equal to 2, which means that electron spin accounts almost entirely, if not entirely, for the observed ferromagnetism. Further evidence for this view is provided by Sucksmith's[85] measurements of the ratio for a series of nickel-copper alloys with low Curie points at temperatures above the latter, using the resonance method designed for his experiments on paramagnetics[86]. The disturbing effects of permanent horizontal magnetic moment of the vertically suspended specimens were great because of iron present as impurity, and the experiments had to be confined to specimens with susceptibilities less than 180×10^{-6}, where the gyromagnetic effect can only be quite small. The value obtained for g was $1{\cdot}9 \pm 0{\cdot}1$ or 2 within the limits of experimental error, under conditions where orbital motion, if present, would profoundly affect the results; thus we may safely conclude that even above the Curie point the magnetic properties are due to spin and not to orbital motion. The discrepancies between Barnett's results and those of other workers, which he attributed to errors arising from magnetostriction effects which occur when the vertical component of the earth's magnetic field is allowed to act on the ferromagnetic during the measurements, have been discussed elsewhere[87].

The value for pyrrhotite is of exceptional interest. Powdered pyrrhotite was

packed inside a fine walled glass tube suspended from a fine quartz fibre inside an evacuated vessel along the axis of a vertical solenoid supplied with current of square wave form. By suitably adjusting the frequency of the current the gyromagnetic effect could be made to produce resonance oscillations of the specimen. The latter carried a mirror which reflected a narrow beam of light upon a horizontal scale in which was cut a narrow vertical slit coinciding with the position of the reflected beam when the specimen was at rest. Light passing through the slit fell upon a photo-electric cell which then activated a relay system and caused the current in the magnetizing solenoid to be reversed. By thus reversing the current at the precise moment that the specimen passed through its equilibrium position many of the disturbing effects which usually arise in gyromagnetic experiments were taken to be eliminated. Inglis[88] has suggested that in pyrrhotite the *d* electrons responsible for ferromagnetism are orientated with their orbital momenta anti-parallel to their spins, i.e. $m_l = 2$ while $m_s = \frac{-1}{2}\, h/2\pi$-units, so that

$$g = (2m_s + m_l)/(m_s + m_l) = 2/3 = 0{\cdot}6\dot{6}.$$

The magnetic behaviour of pyrrhotite is complicated, for it is ferromagnetic in one plane and paramagnetic along directions perpendicular thereto, and as powdered material was used in these experiments, the agreement between experiment and theory is as good as could be expected.

On the view that the elementary particle behaves as a small gyrostat, it would be expected that when a bar of ferromagnetic material is placed with one end in a transverse magnetic field which rotates around the bar, then the latter should be magnetized along its axis in the direction in which a left-handed screw would progress when rotated in the same sense as the magnetic field. Contrary to expectation; Fisher, and more recently Barnett[89], found that the magnetization was extremely small and often in a direction opposite to that required by theory. Auwers[78] has shown, however, that the negative result obtained by experiment follows from the picture of magnetization processes given by Block and Becker, who both suppose that the magnetization of a ferromagnetic takes place, not because of the rotation of whole Weiss domains, i.e. small portions already very strongly magnetized, but merely by changes at the boundaries of such domains. These changes may not necessarily follow the direction of rotation of the field in Fisher's experiment. Indeed, Auwers shows that while this experiment cannot succeed with weak fields, it should give a positive result if rotating fields of 10,000 oersteds could be applied.

Sucksmith[90] has now brought his work on the gyromagnetic ratios of paramagnetics to a close and the values of *g* for ions of the iron and rare earth groups are given in table 5. They were obtained by a series of extraordinarily delicate

Table 5. Values of *g* for paramagnetics

Ion	Cr^{+++}	Mn^{++}	Fe^{++}	Co^{++}	Nd^{+++}	Eu^{+++}	Gd^{+++}	Dy^{+++}
g	1·95	1·98	1·89	1·54	0·78	>4·5	2·12	1·36

resonance measurements with specimens of finely powdered materials packed inside very thin walled glass tubes.

The values for the ions of the iron group are consistent with Stoner's view that their orbital moments are wholly or partially suppressed. In the case of the rare earth group the Van Vleck[91] theory of the magnetic behaviour of these ions is completely verified; as an example we may take Eu^{+++}, for which the earlier Hund theory gave a diamagnetic ion with $g=0$, whereas the Van Vleck value was 6·56. A more precise statement than $g>4{\cdot}5$ could not be given because of the great experimental difficulties attendant upon high values of g.

References

(78) Auwers, O. v. *Naturwissenschaften*, **13**, 202 (1935).
(79) Barnett, S. J. *Rev. Mod. Phys.* **7**, 129 (1935).
(80) Chattock, A. P. and Bates, L. F. *Philos. Trans.* A, **223**, 257 (1923).
(81) Sucksmith, W. and Bates, L. F. *Proc. roy. Soc.* A, **104**, 499 (1923).
(82) Coeterier, F. and Scherrer, P. *Helv. phys. Acta*, **5**, 217 (1932).
Coeterier, F. *Helv. phys. Acta*, **6**, 483 (1933) and **8**, 522 (1935).
(83) Ray-Chaudhuri, D. P. *Nature*, Lond., **130**, 891 (1932).
(84) Sucksmith, W. *Proc. roy. Soc.* A, **108**, 638 (1925).
(85) Sucksmith, W. *Helv. phys. Acta*, **8**, 205 (1935).
(86) Sucksmith, W. *Proc. roy. Soc.* A, **128**, 276 (1930); A, **133**, 179 (1931); A, **135**, 276 (1932).
(87) Bates, L. F. *Nature*, Lond., **134**, 50 (1934).
(88) Inglis, D. R. *Phys. Rev.* **45**, 118 (1934).
(89) Barnett, S. J. *Proc. Amer. Acad. Arts Sci.* **68**, 229 (1933).
(90) Sucksmith, W. *Proc. roy. Soc.* A, **135**, 276 (1932).
(91) Vleck, Van. See discussion by Stoner, *Magnetism and Matter*, p. 309.

§ 7. FERROMAGNETISM

Recent work on alloys. The interesting discovery that metallic gadolinium is ferromagnetic was made by Urbain, Weiss and Trombe[92], using a quantity weighing about 0·12 g. isolated by Trombe[93] from the chloride, which contained less than 0·7 and 0·03 per cent of silicon and iron respectively. Its ferromagnetic Curie point is at $16\pm2^\circ$ C. It is very hard magnetically, although it exhibits a greater specific magnetization than iron at low temperatures, its mean atomic moment being about 3·2 times that of iron. The question of the ferromagnetism of manganese has been tackled by Bates and Reddi Pantulu[94], who have shown that manganese carefully prepared from manganese amalgams is definitely paramagnetic only; although, in conjunction with Gibbs[95] it has also been shown that this manganese readily combines with nitrogen at relatively low temperatures with the formation of ferromagnetic compounds.

A great deal of work has recently been done on the saturation magnetization of ferromagnetics. If the magnetic moment per gram, or specific magnetization of a ferromagnetic, when under the influence of a strong field H and at a temperature T, is denoted by $\sigma_{H,T}$, it is found that, approximately,

$$\sigma_{H,T}=\sigma_{\infty,T}\left(1-\frac{a}{H}\right) \qquad \text{.....(1).}$$

In many cases, a is a constant independent of T, termed the coefficient of magnetic hardness, and $\sigma_{\infty,T}$ represents the saturation specific magnetization in an infinitely strong field at the temperature T; hence the latter quantity can be found by calculation or by graphic means. If, now, we wish to calculate the maximum magnetic moment that the substance can attain, or the saturation specific magnetization at absolute zero $\sigma_{\infty,0}$, then we can make use of the experimental results of Weiss and Forrer[96] which give

$$\sigma_{\infty,T}=\sigma_{\infty,0}\,(1-A.T^2) \qquad \ldots\ldots(2),$$

where A is a constant. The theoretical equation derived by Block[97] on the basis of Heisenberg's[97] exchange theory of magnetism, gives

$$\sigma_{\infty,T}=\sigma_{\infty,0}\,\{1-C.T^{\frac{3}{2}}\} \qquad \ldots\ldots(3).$$

Equation (2) was based upon measurements made over a temperature range down to that of liquid air, and equations (2) and (3) have recently been examined by Fallot[98] in light of experiments extended to liquid-hydrogen temperatures. In Weiss and Forrer's experiments an ellipsoid of the ferromagnetic material under investigation was maintained at one of a series of constant temperatures T in a bath of cooled pentane between the poles of an electromagnet giving a field H after correction for demagnetizing effects of the specimen. The ellipsoid was suddenly removed from the field, and its magnetization measured by a set of coils mounted in the gap and connected in series with a ballistic galvanometer. The coils were designed to eliminate fluctuations of the field, and the effects of the poles or images induced by the magnetized ellipsoid in the pole-pieces of the electromagnet were also found by special experiments. In Fallot's experiments a liquid-hydrogen apparatus of the Ruhemann[99] type was built inside a Dewar vessel. Liquid hydrogen was condensed in a chamber to which the ellipsoid specimen was attached. When the hydrogen evaporated the temperature rose slowly and the necessary measurements could be made by rotating the Dewar flask about an axis of rotation parallel to the line of centres of the pole-pieces, so that the specimen was removed from the field gap. The values of $\sigma_{\infty,T}$ were then plotted against $T^{\frac{3}{2}}$ and T^2, and it was found that equation (3) represented the experimental points better at low temperatures than equation (2), but if extrapolation for $\sigma_{\infty,0}$ had to be made from experimental results for high temperatures only it was better to use equation (2), particularly in the case of nickel; for nickel and nickel alloys exhibit a phenomenon, known as parasitic magnetization[100], of complicated paramagnetic origin. The whole question of the experimental determination of $\sigma_{\infty,0}$ should, however, be re-examined in the light of Stoner's work, described in the following section.

From the quantity $\sigma_{\infty,0}$ the magnetic moment of an individual atom may be found by dividing by Z, the number of atoms in 1 g. The saturation magnetization at absolute zero has been determined for many alloys in order to solve two problems, viz., does the Weiss internal field in a ferromagnetic act upon an atom of another metal dissolved in it to produce a magnetic moment characteristic of the dissolved atom, and, is the atomic moment of the ferromagnetic atom in any way affected by the presence of foreign atoms? The first question was tackled by Sadron[100] and the second by Fallot[98], using the methods already described.

Thus Sadron determined the mean atomic moment of a series of alloys of

different concentrations of chromium, molybdenum and tungsten dissolved in nickel, and deduced that in these circumstances the metals possessed atomic moments of $-3{\cdot}8$, $+3{\cdot}8$ and $+3{\cdot}8$ Bohr magnetons respectively. Such deduction is made by plotting the mean atomic moment against atomic concentration and producing the initial portion of the graph corresponding to low concentrations of the dissolved metal to cut the ordinate at 100 per cent concentration. Manganese dissolved in nickel and cobalt as if it had atomic moments of $+3$ and -3 respectively, and, apparently, in iron with $+0{\cdot}2$; iron was not considered a suitable solvent. The $+$ and $-$ signs in the case of manganese are considered evidence that the Mn atom may set with its magnetic moment parallel or antiparallel to the internal field in the ferromagnetic. From his own and other results Sadron deduced two rules: (I) *The atomic moments of metals in the same column of the periodic table when dissolved in nickel or cobalt are equal*, and (II) *The atomic moment of a metal in the same row of the periodic table as iron varies from element to element in a manner similar to that in which the moment of the corresponding ion in aqueous solution varies.* From a survey of the available evidence it would appear that the above are useful working rules rather than accurate relations.

Although iron was discarded by Sadron, it was used by Fallót, who found the mean atomic moments and Curie points for a number of alloys. Fallot found that in weak solution in iron, silicon did not affect the iron atom, but appeared to reduce its magnetic moment in higher concentrations. Aluminium and chromium atoms appear to reduce the atomic moment of the iron atom, and atoms of gold and tin to increase it. Fallot used a self-recording method for finding the ferromagnetic Curie points of his alloys; other methods for such purpose have recently been described by Forrer and Marian[101], Ochsenfeld[102] and Bates, Gibbs and Reddi Pantulu[93].

The above results on alloys have recently been neatly explained by Mott[103] on the basis of a simple model, in which is postulated a band of energy states corresponding to the $3d$ states of the atom, and another band corresponding to $4s$. In the nickel atom these bands overlap so that there may be p electrons per atom in the s band and p gaps in the d band, where p is about $0{\cdot}6$ per atom. For cobalt and iron p is $0{\cdot}7$ and $0{\cdot}2$ respectively. Mott considers that the maximum binding energy is obtained in these alloys when the number of electrons in the $4s$ band is likewise about $0{\cdot}6$ per atom, and that if extra electrons are introduced by the addition of an atom of copper, zinc, aluminium or tin they must go into the $3d$ band, as long as there is room for them. Here they can only take up states with spins oppositely directed to the direction of magnetization, producing a diminution of magnetic moment by 1, 2, 3 or 4 Bohr magnetons for these metals respectively, in complete agreement with Sadron's results and the early work of Alder[104] on copper-nickel alloys. Now, the manganese atom has three less electrons than nickel; hence, the addition of a manganese atom will decrease the number of electrons in the $3d$ band by three, thus producing an increase in magnetic moment of 3 Bohr magnetons; but no explanation can be given for the decrease found with cobalt, although the experimental results are meagre and it is not certain that the alloy specimens so far used were solid solutions. Again, the addition of palladium

to nickel produces no change in the saturation moment, presumably because both atoms have about 0·6 electrons per atom in the *s* band.

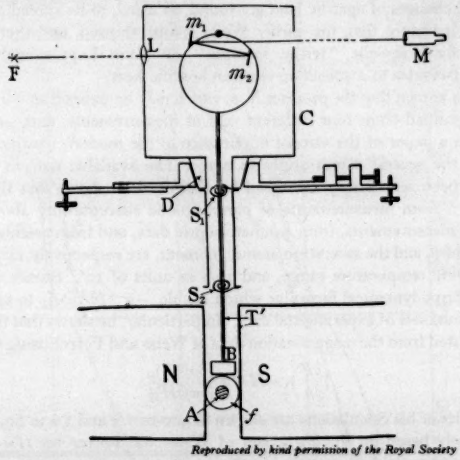
The internal field of ferromagnetics. When we expose a ferromagnetic body to the action of an external magnetic field, the magnetization it acquires is in general dependent both on the magnitude of the applied field and the shape of the body. It may be assumed that the effects of the shape of the body can be eliminated by correction for the demagnetization factor, and H will be taken to denote the field really effective in producing magnetization. According to the Weiss theory, the body in the absence of an external field is built up of a large number of elements or domains each possessing a specific magnetization $\sigma_{0,T}$, the spontaneous, or intrinsic magnetization at the temperature T, when $H=0$. Consequently, a field H as it is gradually increased from zero may do two main things. First, it may align the magnetization vectors of the individual domains, initially by a reversal process in which each vector reverses or turns exactly through 180°, and later by an orientation or turning process in which the vector turns through an angle less than 180°; so that when the vectors are all pointing in the same direction, technical saturation is reached. Second, when H is sufficiently large it may increase the magnetization of each domain to a higher value $\sigma_{H,T}$. Unfortunately, what can be measured by direct experiment is only the apparent magnetization, which we will denote by $\sigma'_{H,T}$, and although this is usually taken to be an exact measure of the intrinsic magnetization $\sigma_{H,T}$ when H is greater than 5000 oersteds, we have no proof that this assumption is correct. Indeed, later in this section evidence will be given to show that it is sometimes seriously in error.

Values of $\sigma'_{0,T}$ are obtained from $\sigma'_{H,T}$ curves by extrapolation to the value of the specific magnetization in infinitely small fields, for temperatures well below the Curie point; again, we have no direct proof that $\sigma'_{0,T}=\sigma_{0,T}$. At higher temperatures, however, this method is unreliable and values of $\sigma'_{0,T}$ are usually derived from magnetocaloric measurements. One of the main objects of the above determinations is, of course, to find how the ratio $\sigma_{0,T}/\sigma_{\infty,0}$ depends upon T/θ_f, since for any ferromagnetic body we expect the ratio to be a unique function of T/θ_f, a statement which is termed the *law of corresponding states*.

The classical experiments of Weiss and Forrer[105] on nickel have formed the basis of many theoretical discussions[106], and our information on ferromagnetics has been greatly increased by Potter's[107] investigations on iron. Potter has obtained accurate values of $\sigma'_{H,T}$ for iron by an ingenious adaptation of the principle of the Kapitza and Webster[108] method of measuring magnetic susceptibilities combined with the use of a Sucksmith ring balance. Potter's arrangements are shown diagrammatically in figure 5. A coil of wire B wound on an ivory former is suspended by a quartz tube T' from the Sucksmith ring balance enclosed in the case C. The coil is mounted with its axis strictly parallel to the uniform field of the magnet NS, so that no axial force arises when a current is passed through B. If an iron sphere A is now placed in the field in the position shown, it becomes magnetized and a vertical force will be set up between A and B when the latter carries a current. This produces a minute axial displacement which is measured by the telemicroscope M, and provides a direct measure of the magnetic moment of A.

The latter is enclosed in a suitable furnace and the required $\sigma'_{H,T}$ measurements can be made without removing A from the field. The flat spiral springs S_1 and S_2 prevent lateral motion.

The evaluation of $\sigma'_{0,T}$ when T lies close to θ_f is made from magnetocaloric measurements. When a piece of a ferromagnetic is suddenly placed in a strong magnetic field, its intrinsic magnetization is increased, the potential energy associated with the Weiss domains is decreased, and, if the energy thus set free cannot leave the material, its temperature must rise. Similarly, if the material is suddenly removed from the field a fall in temperature must result. This magnetocaloric



Reproduced by kind permission of the Royal Society

Figure 5. Potter's apparatus for measurement of the magnetocaloric effect in iron.

phenomenon, apart from minute hysteresis effects, is strictly reversible, and thermodynamical reasoning shows that the change in temperature ΔT is equal to

$$\frac{N}{2J.\rho.C_\sigma}(\sigma^2_{H,T}-\sigma^2_{0,T})$$

when $T<\theta_f$, and to

$$(N+1/\chi)\frac{\sigma^2_{H,T}}{2J.\rho.C_\sigma}=A.\sigma^2_{H,T},$$

when $T>\theta_f$, N being the constant of the Weiss internal field, ρ the density, C_σ the specific heat at constant magnetization, J the mechanical equivalent and χ the paramagnetic susceptibility. Consequently, if we plot ΔT against the corresponding values of $\sigma^2_{H,T}$ for T constant we should get a linear graph whose intercept should give $\sigma^2_{0,T}$ at the chosen temperature T. In practice we measure $(\sigma'_{H,T})^2$ and the graphs exhibit flat feet, but definite extrapolations can be made from the values with

strong fields. In Potter's experiments the values of ΔT were measured by a thermo-junction embedded in the sphere A.

On the Weiss theory the energy per unit mass associated with the intrinsic magnetization of a ferromagnetic is $E=-\frac{1}{2}N.\rho.\sigma^2_{H,T}$, and the specific heat of a ferromagnetic in zero field must therefore exceed that of a normal metal by an amount

$$-\frac{N.\rho}{2J}\frac{d(\sigma_{0,T})^2}{dT}.$$

A new set of specific-heat measurements on nickel have recently been made by Grew[109], using methods similar to those originally employed by Mme Lapp[110]. The marked changes of specific heat are found, as usual, to be spread over a wider range of temperature than the earlier Weiss results showed, and there is a small excess specific heat—the "terme inconnu" in Mme Lapp's analysis—which Stoner[111] attributes to a collective electron specific heat.

Now, it is known that the product $N.\rho$, which may be defined as $-\sigma^{-1}(\partial E/\partial\sigma)_T$, can be determined from four different sets of measurements, and, as Stoner[112] points out in a paper of the utmost importance to the modern treatment of ferromagnetism, the several values ought to agree. The available data in the case of nickel have been searchingly examined by Stoner, who shows that the values of $N.\rho$ obtained from measurements of paramagnetic susceptibility above θ_f, from specific heat measurements, from magnetocaloric data, and from calculations based on the value of θ_f and the saturation atomic moment, are respectively 11·7, 11·5, 4·3–9·8 over a wide temperature range, and 16·4 in units of 10^4. Stoner discusses in detail the thermodynamical formulae which enable $-\sigma^{-1}(\partial E/\partial\sigma)_T$ to be calculated from the various sets of experimental data. In particular, he shows that this quantity can be calculated from the magnetization data of Weiss and Forrer using the formula

$$\left(\frac{\partial E}{\partial\sigma}\right)_T=H+T\frac{(\partial\sigma/\partial T)_H}{(\partial\sigma/\partial H)_T}.$$

The results of his calculations are shown in curves 1 *a* and 1 *b* in figure 6 which are respectively based on the $\sigma'_{H,T}$ data of Weiss and Forrer for $H=15{,}000$ and $H=7000$. A very marked variation of $N.\rho$ with temperature is thus brought to light, and it is obvious that an increase occurs up to temperatures well above the Curie point.

Now, Potter first drew attention to the fact that a variation of $N.\rho$ with temperature could be deduced from Weiss and Forrer's magnetocaloric measurements on nickel, and from the same data Stoner obtains the curves 2 *a* and 2 *b*, which again refer to $H=15{,}000$ and $H=7000$ respectively. The two sets of curves are clearly of the same configuration although the magnetocaloric values of $N.\rho$ appear definitely the lower and have to be corrected for an experimental error, which brings the values from the two sets of data into quite close agreement. In both cases, however, the values of $N.\rho$ are greater for the higher fields.

In curve 3 of figure 6 are reproduced the values of $N.\rho$ calculated from the specific-heat data of Lapp and Grew. Now the values of $N.\rho$ shown in curve 3 depend only on the spontaneous magnetization $\sigma_{0,T}$, whereas the values of curves 1 and 2, from the magnetization and magnetocaloric data, depend either directly or

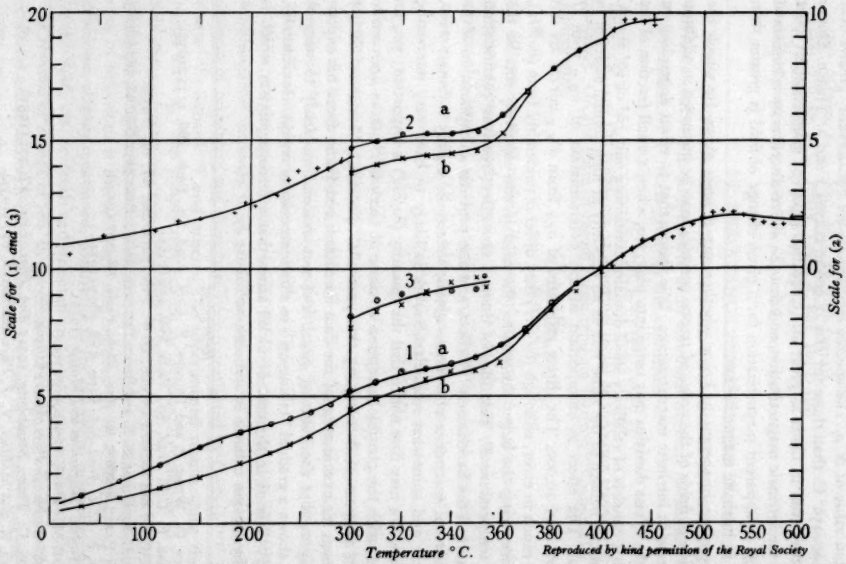


Figure 6. Curves showing variation of $N.\rho$ with temperature (Stoner). Nickel. $-(1/\sigma)(\partial E/\partial\sigma)_T \times 10^{-4}$. Curve 1, from magnetization data. Curve 2, from magnetocaloric data. (*a*) $H=15{,}000$; (*b*) $H=7000$. The values are independent of H above 400° C. For (2) below 300° C., $H=12{,}000$. Curve 3, from specific-heat results. ⊙ Lapp; × Grew. For curves 1 and 2, σ = apparent magnetization. For curve 3, σ = extrapolated int insic magnetization for $H=0$. Note change of scale at $T=300$ and $T=400$.

indirectly upon the quantity $\partial(\sigma_{H,T})/\partial H$, and, the greater this quantity, the smaller the calculated value of $N.\rho$. The specific heat results are considerably greater over the range 300–350° C. than those for $H=15{,}000$ in curves 1 and 2. Hence, Stoner concludes that even in fields as great as 15,000 the apparent magnetization must differ from the intrinsic magnetization, and that the above discrepancy is explained if the change in apparent magnetization for a given change of field is greater than the change in intrinsic magnetization.

It is therefore important to know whether any process exists by which the magnetization in fields of the order of 10,000 oersteds may be increased in addition to the increase in intrinsic magnetization. It is shown that an effect depending on the size of the Weiss domains may come into play; for, when a small fraction of the ferromagnetic consists of relatively small domains containing only 10^3 or 10^4 atoms, this fraction can behave in the same way as an impurity of high atomic moment in the material. The effect is unimportant in the determination of $\sigma_{H,T}$ with H between 5000 and 20,000. The direct estimates of $\sigma_{0,T}$ from $\sigma'_{H,T}$ curves are not likely to be much in error, although $(\partial\sigma'/\partial H)$ may differ considerably from $(\partial\sigma/\partial H)$, the difference being the larger the smaller the field H and relatively smaller the higher the temperature T. This idea of domain size effect receives confirmation from the study of the Barkhausen[105] effect, and also provides an explanation of the fact that different specimens of a ferromagnetic behave in slightly different ways. Stoner's new and somewhat iconoclastic work is likely to have very interesting repercussions. We may first notice that the quantity $(\partial\sigma/\partial H)$ is important; yet, no experiments, with the possible exception of those of Gerloff[113], have been made with the set purpose of determining this particular quantity. Moreover, Stoner considers from his examination of the data for nickel, and Potter from his experiments on iron, that the ratio $\sigma_{0,T}/\sigma_{\infty,0}$ does not descend suddenly to zero at $T=\theta_f$, but shows a gradual "tailing off" in this temperature region. Incidentally, Frl. Scharff[114] finds in the case of nickel wires that a ferromagnetism which is influenced by tension extends to some 12–20° above θ_f.

References

(92) Urbain, G., Weiss, P. and Trombe, F. *C.R. Acad. Sci.*, Paris, **200**, 2132 (1935).
(93) Trombe, F. *C.R. Acad. Sci.*, Paris, **200**, 459 (1935).
(94) Bates, L. F. and Reddi Pantulu, D. V. *Proc. phys. Soc.* **47**, 197 (1935).
(95) Bates, L. F., Gibbs, R. E. and Reddi Pantulu, D. V. *Proc. phys. Soc.* **48**, 665 (1936).
(96) Weiss, P. and Forrer, R. *Ann. Phys.*, Paris, **12**, 297 (1929).
Weiss, P. *C.R. Acad. Sci.*, Paris, **198**, 1893 (1934).
(97) See Stoner. *Magnetism and Matter* (1935).
(98) Fallot, M. Thèse, Strasbourg (1935).
(99) Ruhemann, M. *Z. Phys.* **65**, 67 (1930).
(100) Sadron, C. Thèse, Strasbourg, 1932; *Ann. Phys.*, Lpz., **17**, 371 (1932).
(101) Forrer, R. and Marian, V. *J. Phys. Radium*, **5**, 617 (1934).
(102) Ochsenfeld, R. *Ann. Phys.*, Lpz., **12**, 353 (1932).
(103) Mott, N. F. *Proc. phys. Soc.* **47**, 571 (1935).
(104) Alder. Thesis, Zürich (1916).
(105) Weiss, P. and Forrer, R. *Ann. Phys.*, Paris, **5**, 153 (1926).
(106) See Stoner. *Magnetism and Matter.*
(107) Potter, H. H. *Proc. roy. Soc.* A, **146**, 362 (1934).
(108) Kapitza, P. and Webster, W. L. *Proc. roy. Soc.* A, **132**, 442 (1931).

(109) GREW, K. E. *Proc. roy. Soc.* A, **145**, 509 (1934).
(110) LAPP, Mme. *Ann. Phys.*, Paris, **12**, 442 (1929).
(111) STONER, E. C. *Phil. Mag.* **15**, 1018 (1933).
(112) STONER, E. C. *Philos. Trans.* A, **235**, 165 (1936).
(113) GERLOFF, G. *Z. Phys.* **99**, 585 (1936).
(114) SCHARFF, G. *Ann. Phys.*, Lpz., **25**, 223 (1936).

§ 8. RESISTANCE MEASUREMENTS

It has been known for many years that the resistance of a ferromagnetic metal increases with rise in temperature in an abnormal manner, in fact its resistance R_T at a temperature $T°$ K. is given by the expression

$$R_T = (R_n)_T + c\,(\sigma^2_{\infty,0} - \sigma^2_{0,T}),$$

where $(R_n)_T$ is the resistance of a normal metal at the same temperature and c is a constant. On the basis of recent measurements Gerlach[115] concludes that at any temperature

$$\frac{(R_n)_T - R_T}{R_0} = b\,.\,\sigma^2_{0,T}$$

for all specimens of nickel, R_0 being the resistance of the nickel at 0° C. and b a constant approximately equal to 5×10^{-4}. He further concludes, from the experiments of Löwe[116] on the thermal emissivity of nickel and the Hagen-Rubens relation $(\epsilon)_\lambda = 0{\cdot}0365\,\sqrt{(R_T/\lambda)}$, where λ denotes the wave-length of the radiation, that

$$\frac{(\epsilon_n)^2_T - (\epsilon)^2_T}{\epsilon_0{}^2} = b\,.\,\sigma^2_{0,T},$$

where $(\epsilon_n)_T$ is the emissivity of a normal metal and $(\epsilon)_T$ that of nickel at $T°$, while (ϵ_0) is the emissivity at 0° C. It is very interesting that the latter relation is obeyed within the limits of experimental error for wave-lengths of 8·7 and 24 μ. Bates and Hogwood[125] had earlier attempted to detect Raman lines in the light scattered from a ferromagnetic oxide, but without success.

Now, on the application of a field H, the intrinsic specific magnetization of a ferromagnetic increases to $\sigma_{H,T}$. Consequently, we should expect corresponding changes of resistance, and these have been investigated by Gerlach[115], Englert[117], Potter[118], Scharff[119] and others. The experimental results show that iron and nickel wires at room temperature exhibit an increase in resistance in a longitudinal field and a decrease in a transverse field as shown in figure 7. At temperatures near the Curie point decreases are observed in both fields. Gerlach explains this difference in behaviour at low and high temperatures on the view that the two distinct processes, already mentioned, contribute to the change in magnetization produced by the field, viz. reversals and orientations (Umklappprozesse and Drehprozesse). The reversal or change of 180° in direction of the magnetic vector of a Weiss domain produces a comparatively small change, a decrease of resistance, but the orientation of the vector of the domain (or extension of the boundary on the Becker view) causes a comparatively large change, an increase of resistance in longitudinal fields. The latter process is very important at low temperatures.

It is well known that the thermoelectric power of a ferromagnetic exhibits an anomaly in the neighbourhood of the ferromagnetic Curie point, and several unsuccessful attempts have been made to correlate changes in the Thomson coefficient

with the magnetic changes. There has recently been a lull in such measurements, but Foster[120] has shown that the addition of only 1 and 2 per cent of chromium to nickel decreases the change in the Thomson coefficient at the Curie point of the alloy to only 30 and 10 per cent that of pure nickel, respectively.

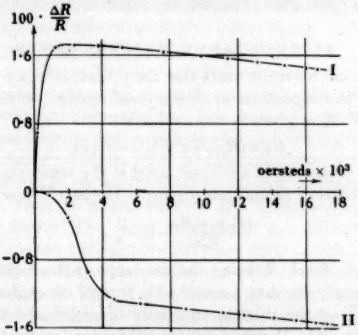


Figure 7. Changes of resistance of nickel at room temperature in a magnetic field, I Longitudinal, II Transverse.

Schulze[121] has examined the magnetization and magnetostriction data for iron and nickel single crystals and for polycrystalline nickel and finds that

$$\Delta l/l = c\,(\sigma^2_{0,T} - \sigma^2_{\infty,0}).$$

Magnetic methods of producing very low temperatures were recently described by McLennan[122], and a very clear account of special methods and applications is given by Simon[123], and by Kürti and Simon[124].

References

(115) Gerlach, W. *Ann. Phys.*, Lpz., **25**, 210 (1936).
(116) Löwe, E. *Ann. Phys.*, Lpz., **25**, 210 (1936).
(117) Englert, E. *Ann. Phys.*, Lpz., **14**, 589 (1932).
(118) Potter, H. H. *Proc. roy. Soc.* A, **132**, 560 (1931); *Phil. Mag.* **13**, 233 (1932).
(119) Scharff, G. *Z. Phys.* **97**, 73 (1935).
(120) Foster, A. W. *Phil. Mag.* **18**, 470 (1935).
(121) Schulze, A. *Z. Phys.* **82**, 674 (1933).
(122) McLennan, J. C. *Reports on Progress in Physics*, **1**, 204 (1934).
(123) Simon, F. *Not. Proc. roy. Instn* (1935).
(124) Kürti, N. and Simon, F. *Proc. roy. Soc.* A, **149**, 152 (1935).
(125) Bates, L. F. and Hogwood, H. E. *Proc. phys. Soc.* **47**, 878 (1935).

§ 9. MAGNETOSTRICTION PHENOMENA

A considerable number of papers has been published during the last fifty years on the mechanical strains which accompany magnetization and the magnetic changes produced by mechanical stresses. The production of single crystals of ferromagnetics and the importance of the magnetostriction oscillator in navigation, together with a

considerable advance in the theoretical treatment, have aroused fresh interest in the phenomena, although experimental methods remain much the same.

The main advance in the theoretical treatment is due to Becker[126]. He starts with the familiar Weiss picture of a ferromagnetic consisting of magnetized domains, and considers the total energy of a distorted dipole lattice under the influence of an applied field to be made up of the internal magnetic energy of the lattice, the energy stored in the elastic deformation and the potential energy of the domains in the applied field. The I, H curve can then be obtained from the condition that the total energy shall be a minimum, and a fairly satisfactory, though qualitative, picture of technical hysteresis cycles is thus provided.

Now, it is well known that when a nickel wire is magnetized it shortens, and Becker therefore deduced that, when under the simultaneous action of a longitudinal field H and a severe longitudinal stress Z, the wire should be magnetized in such a way that the magnetic vectors make an angle α with the axis of the wire given by

$$\alpha = \cos^{-1} \frac{I_\infty}{3\lambda_c . Z} H \qquad \ldots\ldots(1),$$

where I_∞ is the technical saturation intensity of magnetization and λ_c the experimental value of the magnetostrictive contraction of the saturated wire. It therefore followed that the component of magnetization parallel to the axis was

$$I_\parallel = \frac{I_\infty{}^2}{3\lambda_c . Z} H = \frac{C . H}{Z} \qquad \ldots\ldots(2),$$

where C is a constant. From equation (1) we see that a severely strained wire, in the absence of a magnetic field, should have the magnetic vectors set radially. From equation (2) we see that a linear $\{I_\parallel . H\}$ curve can be obtained by subjecting a nickel wire to severe tension, and that the field necessary to produce technical saturation varies directly with the tension. It can further be shown that the magnetostrictive contraction of a nickel wire under severe tension should be 1½ times that of an unstretched wire. These deductions have been confirmed by the experiments of Becker and Kersten[127] using a form of apparatus designed by Preisach[128]; observations of Barkhausen effects were also made at the same time with Pfaffenberger's[129] arrangements.

Now, the effects of the shape of the specimens used in magnetostriction experiments on the results have only recently been studied. In a theoretical study Becker[130] showed that the lateral and longitudinal changes in dimensions of an ellipsoid placed in a weak magnetic field were respectively greater by $\frac{1}{2}I^2 . D\ (1/3k - \alpha'/4\eta)$ and $\frac{1}{2}I^2 . D\ (1/3k + \alpha'/2\eta)$, and the volume change greater by $\frac{1}{2}I^2 . D/k$ for the same intensity of magnetization I, than for an infinitely long and thin rod; k is the compressibility, η the modulus of rigidity, D the demagnetization factor of the ellipsoid and α' a constant which represents the effects of strains upon D. Since in weak fields $H = D . I$, the volume change is greater by an amount $\frac{1}{2}H^2/kD$.

Assuming that in strong fields the magnetization acquired is affected only by changes in the volume V, so that $V . I = V_0 . I_0\ (1 + e . w)$, where w is the volume increase and e a constant, it can be shown that the magnetostriction change in volume increases linearly with H. The curve of volume change plotted as a function of

($H-D.I$) should therefore start with a parabolic portion and end with a linear portion, these being joined by an intermediate portion. The initial portion is determined by the dimension ratio of the ellipsoid, the final portion by the variation of the intrinsic magnetization with V, and the intermediate region by the turning of the magnetic vectors of the Weiss domains.

An experimental study of the effect of shape on longitudinal measurements was made by Kornetzki[131] who used five iron ellipsoids of revolution each 10 mm. long with axis ratios a/b respectively equal to 1, 2, 3, 4 and 5, and who found that the mean experimental value of the magnetostriction coefficient $\Delta l/l$ for a sphere was $-3{\cdot}15\times10^{-6}\pm5$ per cent, compared with the experimental value -7×10^{-6} obtained for a very long rod, the calculated value for the sphere being $-2{\cdot}8\times10^{-6}$. The volume changes of iron ellipsoids were also investigated by Kornetzki[132], the experimental procedure being based on that of Honda and Nagaoka[133], and Becker's deductions were confirmed in a remarkable degree.

From the final linear increase in volume with magnetic field Kornetzki has attempted to obtain information concerning the behaviour of the Weiss constant N, or, more precisely, $N.\rho$, and he discusses the change in the ferromagnetic Curie point with pressure. In extreme cases, he suggests, the latter may actually be depressed below room temperature, and this may be the explanation of Steinberger's[134] observation that a 30 per cent nickel-iron alloy loses its ferromagnetic properties when exposed to a pressure of 10,000 atmospheres.

Strains which must influence magnetization may be set up in any ferromagnetic in three main ways. First, there may be present foreign atoms of non-metallic nature, which do not fit into the regular crystal lattice of the pure material. An excellent example of the effect of the removal of such strains is provided in the case of hydrogenized iron, from which these impurities have been removed by heating the metal in a stream of moist hydrogen at high temperatures[135]. Second, they may be produced by mechanical means, such as cold working, and Bozorth[136] refers to some unpublished data by Dillinger and Haworth on the increase in permeability of permalloy tapes as the annealing temperature is raised following cold rolling, in which X-ray examination shows that there is a progressive reduction in internal strains as the annealing temperature is raised. Internal strains are very marked in permanent magnet materials which are hardened by the precipitation process[137], in which one phase of an alloy is made to separate out in a finely divided form throughout the material. This is done by keeping the alloy, which has previously been quenched from a temperature at which it was a solid solution, at a temperature at which precipitation slowly takes place, and finally cooling it when the process is sufficiently complete. Third, strains may be set up by magnetostriction processes, and Dillinger and Bozorth[138] have shown how these may successfully be used to favour magnetization in a particular direction, simply by annealing the alloy at a suitable temperature slightly above its Curie point in a magnetic field of moderate strength; for example, the maximum permeability can be greatly increased in this way, and the material made to exhibit a hysteresis cycle of rectangular shape. High permeability in one direction can only be achieved with loss of permeability in others, and in some cases the alloy is given a maximum permeability

150 times as great in one direction as in another at right angles. Dillinger and Bozorth[139] give a simple picture of the processes involved and point out the similarity between them and those involved in the formation of superstructures investigated mathematically by Bragg and Williams[140].

The presence of domains in ferromagnetics is strikingly illustrated by the formation of regular patterns[141] deposited on a ferromagnetic when its surface is covered with a colloidal solution of iron oxide, even when the metal is demagnetized, and, on the whole, magnetostriction phenomena conclusively support the Weiss domain picture of ferromagnetism. Of the average size of the domains the Barkhausen phenomena give some information[142], and in view of Stoner's work on domain size effect there is need for more detailed knowledge of domain sizes.

The importance of magnetostriction oscillators is brought out in the excellent work of Wood, Smith and McGeachy[143] on a magnetostriction echo depth recorder. By means of suitable reflectors magnetostriction oscillations are directed towards the sea-bed from the bottom of a boat. The echo generates high-frequency currents in a neighbouring receiver, and these are amplified and passed through a chemical recorder. The motor driving the latter controls the instant of transmission of the sound signal, so that the motion of the recording stylus is synchronized with the regular sequence of oscillation signals. The stylus produces two marks on a recording drum, one corresponding to the transmission, and the other to the reception of the echo, and as the ship travels on its course a continuous record of the depth of the water is provided. The instrument measures to an accuracy of about 1 ft. in depths from zero to 30 fathoms, and, clearly, is likely to be widely used.

References

(126) Becker, R. *Z. Phys.* **62**, 253 (1930) and **64**, 660 (1930). Cf. Müller-Pouillet, *Lehrbuch der Physik*, **4**, Pt. 4, 786 (1934).
(127) Becker, R. and Kersten, M. *Z. Phys.* **64**, 660 (1930).
(128) Preisach, F. *Ann. Phys.*, Lpz., **3**, 737 (1929).
(129) Pfaffenberger. *Ann. Phys.*, Lpz., **87**, 737 (1928).
(130) Becker, R. *Z. Phys.*, Lpz., **87**, 547 (1933); see also Hayasi, T. *Z. Phys.* **72**, 177 (1931).
(131) Kornetzki, M. *Z. Phys.* **87**, 560 (1933).
(132) Kornetzki, M. *Z. Phys.* **98**, 289 (1935).
(133) Honda, K. and Nagaoka, H. *Phil. Mag.* **4**, 56 (1902).
(134) Steinberger, R. L. *Physics*, **4**, 154 (1933).
(135) Coiffi, P. P. *Phys. Rev.* **45**, 742 (1934).
(136) Bozorth, R. M. *Elect. Engng*, **54**, 1251 (1935) and *Bell Syst. tech. J.* **15**, 63 (1936).
(137) Köster, W. *Stahl u. Eisen*, **53**, 849 (1933).
(138) Dillinger, Joy F. and Bozorth, R. M. *Physics*, **6**, 279 (1935).
(139) Dillinger, Joy F. and Bozorth, R. M. *Physics*, **6**, 285 (1935).
(140) Bragg, W. L. and Williams, E. J. *Proc. roy. Soc.* A, **145**, 699 (1934).
(141) Bitter, F. *Phys. Rev.* **41**, 507 (1932).
McKeehan, L. W. and Elmore, W. C. *Phys. Rev.* **46**, 226 (1934).
Snoek, J. L. *Physica*, 's Grav., **3**, 118 (1936).
(142) Bozorth, R. M. and Dillinger, Joy F. *Phys. Rev.* **35**, 733 (1930).
(143) Wood, A. B., Smith, F. D. and McGeachy, J. A. *Proc. Instn Elect. Engrs*, **76**, 550 (1935).

§ 10. ACKNOWLEDGMENT

The writer has had the pleasure of discussing many sections of this report with Dr W. Sucksmith of the H. H. Wills Laboratory, Bristol.

EXPERIMENTAL ELECTRICITY AND MAGNETISM

§ 1. DIELECTRICS

By L. HARTSHORN, D.Sc., A.R.C.S., D.I.C., A.Inst.P.

IN the fundamental electromagnetic equations on which nearly all electrical and magnetic measurements are based, the electrical properties of any material substance are represented by three constants: the magnetic permeability μ_m, the dielectric constant or permittivity ϵ, and the conductivity σ, and it is often convenient to divide materials into three classes: magnetic materials, dielectrics, and conductors, according as the phenomena under consideration are mainly dependent on μ_m, ϵ, or σ respectively. There is of course no sharp line of demarcation between the classes, and the same substance may fall into any of them, depending on the phenomena under discussion. It is proposed to consider here the trend of some of the recent work on dielectrics, or on those properties of materials which are represented by ϵ, but it will be found that the conductivity σ for such bodies is also largely controlled by the same molecular or ionic processes as determine ϵ, and the two properties must in general be considered together. The conductivity σ for such substances is by no means constant. When the applied voltage is alternating the values of σ obtained from measurements at two different frequencies may not be even of the same order of magnitude, and it is therefore sometimes convenient to represent conductance phenomena by other coefficients. Two of these are widely scattered through the literature, viz. the loss tangent, tan δ (which is usually practically the same as the power factor sin δ, since δ is mostly small), and the absorption index κ. Again, many writers use, instead of the two coefficients ϵ and σ, a single complex coefficient $\epsilon_c = \epsilon' - j\epsilon''$. The significance of these terms may be briefly indicated as follows.

The conductivity σ is a measure of the conductance current flowing in a material under a given voltage gradient, while the dielectric constant ϵ is a measure of the displacement current flowing under a given voltage gradient at a given frequency, this current being also proportional to the frequency and in quadrature with the voltage and the conductance current. The loss tangent tan δ is the ratio of the conductance current to the displacement current flowing under the same conditions. The power dissipated in the material is proportional to $V^2\sigma$ and to $V^2\omega$ tan δ, and δ is the phase difference between the total current and the displacement current. Alternatively, we may regard the whole current as a displacement current and represent the change of phase δ by the use of the complex coefficient $\epsilon_c = \epsilon' - j\epsilon''$. Evidently tan $\delta = \epsilon''/\epsilon'$. Again, experiments at very high frequencies often involve the consideration of electromagnetic waves and it may then be convenient to express the same quantities in terms of the refractive index r (the ratio of the wave-lengths in the material and in free space) and absorption index κ, such that the amplitude of the field strength of a train of plane waves in the medium is reduced, on account

of energy dissipation in the material, by the factor $e^{-2\pi\kappa}$ after the wave has travelled a distance of one wave-length. We may note that

$$\epsilon_c = \epsilon' - j\epsilon'' = r^2 (1 - j\kappa)^2.$$

It will be clear that when a material dissipates energy at a rate which is nearly independent of the frequency of the electric intensity, it will be convenient to express the results in terms of conductivity. When, however, the energy loss increases rapidly with increase of frequency it will be more convenient to consider the power factor, loss tangent, or the absorption index in conjunction with the dielectric constant.

Electric moments. The electromagnetic theory of light indicates that the dielectric constant of a substance in which the dissipation of energy is insignificant should be equal to the square of its refractive index, and until the advent of Debye's theory of polar molecules it was very difficult to account for the fact that many substances such as water and nitrobenzene showed a dielectric constant of an order of magnitude quite different from that of the square of the optical refractive index. These liquids possessed values of ϵ which were very high and which diminished rapidly with rise of temperature. Debye[1] explained these facts by the assumption that the molecules of such materials were more or less rigid yet electrically unsymmetrical, so that they possessed a permanent electric moment, and would therefore, on the application of an electric field, tend to take up a definite orientation in which their electric axes were in line with the field. The material therefore becomes electrically polarized by the field just as a magnetic material is polarized by a magnetic field, and its condition is conveniently represented by the intensity of polarization I or electric moment per unit of volume $(D-E)/4\pi = (\epsilon - 1) E/4\pi$, where D is the electric displacement and E the applied electric force. The mean moment per molecule will be given by $(\epsilon - 1) EM/4\pi\rho N$, where ρ is the density, M is the molecular weight, and N is Avogadro's number. The magnitude of the polarization for any given applied field and therefore the value of ϵ will evidently depend on the forces resisting the orientation of the molecules and in gases and liquids these will certainly be mainly those arising from thermal agitation, which will tend to set up a random distribution of orientations which corresponds to zero mean moment in any one direction. When the field is applied the tendency of the molecules to take up the orientation of the field is increased, and there will be a finite mean moment in this direction of a magnitude which must diminish with rise of temperature. Thus the dielectric constant arising from molecular orientation will certainly diminish with rise of temperature in liquids and gases. If the permanent moment of each molecule is μ and if under the influence of the applied field the orientations are distributed according to Boltzmann's law, then the mean moment is given approximately by $\mu^2 F/3kT$ so long as $\mu F/kT$ is very small. Here T is the absolute temperature, k is Boltzmann's constant, and F is the electric field acting on the molecules. It follows that

$$\frac{(\epsilon - 1) ME}{4\pi N\rho} = \frac{\mu^2 F}{3kT} \qquad \text{.....(1).}$$

It is to be noted that the internal field F differs from the applied field E owing to the electric field of the molecules themselves. In a gas, however, we may assume that the molecules are so far apart that the mean effect on any one molecule of the electric fields of the others is negligible and we may write $F=E$, and deduce that for gases a dielectric constant entirely due to the orientation of polar molecules should obey the law

$$\frac{(\epsilon-1)}{\rho}\frac{M}{4\pi N}=\frac{\mu^2}{3kT} \qquad \text{.....(2),}$$

from which it should be possible to calculate the molecular moment μ. It must be remembered, however, that the electric displacement in any material will not be entirely due to the orientation of molecules, for the field will exert a force on all the ions and electrons constituting the material and there will therefore be a separation of positive and electric charges within the molecule against the action of the elastic forces holding them in their normal positions. This constitutes an additional electrical displacement and of course gives each molecule an additional induced electric moment, which is proportional to the applied field and vanishes when the field is removed. This induced electric moment may be represented by aF, where a is practically independent of temperature. Equation (1) then becomes

$$\frac{(\epsilon-1)}{\rho}\frac{ME}{4\pi N}=\left(\frac{\mu^2}{3kT}+a\right)F \qquad \text{.....(3),}$$

and the constant term a must therefore be added to equation (2). When we attempt to apply the same ideas to liquids and solids we are faced with the difficulty of determining the relation between the applied field E and the internal field F. If we assume that this is such as to satisfy the Clausius-Mosotti hypothesis we have $F=E+4\pi I/3$, and instead of equation (2) we have

$$\frac{(\epsilon-1)}{4\pi\rho N}M=\left(\frac{\mu^2}{3kT}+a\right)\frac{(\epsilon+2)}{3} \qquad \text{.....(4).}$$

Thus the molecular constant $4\pi N(\mu^2/3kT+a)/3=P$ is measured by $(\epsilon-1)\,M/(\epsilon+2)\,\rho$ for any material in which the internal field corresponds to the Clausius-Mosotti hypothesis. It is evident that this quantity, usually called the molar polarization, may be readily determined by measurements of dielectric constant and density. On plotting it against $1/T$ a straight line should be obtained, the slope of which should give the value of μ, the electric moment of the molecule, while the intercept on the axis of P should give a. The equation is frequently written in the form

$$\frac{(\epsilon-1)}{(\epsilon+2)}\frac{M}{\rho}=P=P_E+P_A+\frac{B}{T}=A+\frac{B}{T} \qquad \text{.....(5).}$$

Evidently B/T represents that part of the polarization due to molecular orientation; the remainder A is considered to consist of two parts, P_E which represents that due to the displacements of electrons within the atoms, and P_A that due to the vibrations of the atoms. Luminous radiation will cause displacements of the electrons only, so that P_E may be obtained from the optical refractive index of the material. The

procedure described above determines $A = P_E + P_A$, and hence P_A, the atomic polarization, may be obtained by difference.

In recent years there have been great improvements in the methods for the measurement of dielectric constant. Some of these were briefly recorded in the *Progress Report* for 1934. The technique may now be considered to be well established and it has therefore become a straightforward matter to test the validity of the above equations with almost any material of reasonably small conductivity. The assumptions made are very considerable, and it would certainly not be very surprising if the internal field were not such as to satisfy the Clausius-Mosotti relation. Gases and vapours are more likely to satisfy the conditions than liquids and solids, and therefore results obtained with these are likely to be the most trustworthy. However, measurements on liquids are easier and therefore much of the earlier work was done with them. A very dilute solution of a material with polar molecules in a non-polar solvent has some of the advantages of a gas, and many measurements have been made with such solutions, the equations of course being slightly modified to take into account the existence of the two kinds of molecules in the solution. It would be very surprising if solids were found to satisfy the above assumptions, but it will be found nevertheless that the application of some of the ideas to solids has led to interesting results.

Experiments on gases and vapours. It has now been widely recognized for several years that many substances satisfy equation (5) and that it is possible to assign to them a definite molecular moment. Indeed in 1934 a Table of Dipole Moments[2] was published which gives the values for about a thousand chemical substances, together with a list of some 300 original papers from which they were taken. The output of work of this kind shows no sign of diminution. It appears that a recognition of the advantages of working with gases and vapours has led to the planning of a series of systematic measurements on gases and vapours in several parts of the world. Thus quite recently we have had papers on a series of alcohols by Kubo[3,4], and by Mahanti[5], on various benzene derivatives and alkyl chlorides by Smyth and McAlpine[6,7], and by Mahanti[8], and on a series of vapours by Groves and Sugden[9]. It may be said that all the materials examined in these investigations satisfy equation (5) with precision and there can be little doubt that the values obtained represent the electric moment of free molecules of the materials in question. An idea of the accuracy now available may be obtained by comparing the results of three different observers on the same substance. Thus for the electric moment of the water molecule we have the values

$1{\cdot}842 \pm 0{\cdot}008 \times 10^{-18}$	c.g.s.e. unit	(Sänger)[10],
$1{\cdot}831 \pm 0{\cdot}006$	,,	(Stranathan)[11],
$1{\cdot}84 \pm 0{\cdot}01$	,,	(Groves and Sugden)[9].

When it is remembered that these values depend on measurements of $\epsilon - 1$, which is a very small quantity for a vapour, it will be realized that this agreement represents a considerable achievement. It may be noted here that several experimenters[9, 12, 13] have compared the moments of heavy water and ordinary water and all are agreed

that the values are practically the same. On the other hand, de Bruyne and Smyth[14] find that the value for deutero-ammonia ND_3 exceeds by $0{\cdot}03 \times 10^{-18}$ the value for NH_3, which they find to be $1{\cdot}466 \times 10^{-18}$.

Several interesting generalizations have been deduced from the values for various materials. Thus the fatty hydrocarbons have zero molecular moment, but derivatives containing groups such as OH, NO_2, Cl, Br, or I are polar, the moment depending almost entirely on the substituted group or atom and not on the length of the hydrocarbon chain[3,4,6,9]. It may well become possible ultimately to assign values to the more important group and to calculate the resultant moment of any molecule. The chemical aspects of the problem have recently been summarized in more detail by J. W. Williams[15].

An interesting development on the physical side has been made by Kubo[16]. We have seen that Debye's equation has been found to hold for many gases and vapours at ordinary low pressures, and we may therefore consider the Clausius-Mosotti relation to hold for this condition. The question arises, does the relation still hold when the pressure of the gas is increased to very high values? Kubo argues that in a polar gas each molecule must be polarized to some extent by the electric field of neighbouring molecules, and that the moment induced in this way must be added to the moment of the free molecule in order to get the resultant moment. The induced moment will evidently increase with the moment of neighbouring molecules and also their proximity, i.e. with the value of μ and the pressure or the molecular volume V. Kubo obtains the relation

$$P = \frac{4\pi N\alpha}{3} + \frac{4\pi N\mu^2}{9kT}\left[1 + \frac{2\pi\alpha\mu^2 N}{9kT\beta^3 V}\right],$$

where β is the radius of the sphere of molecular action. Thus for a polar gas at constant temperature, P may be expected to be constant only at low pressures for which the correcting term $2\pi\alpha\mu^2 N/9kT\beta^3 V$, which has been added to Debye's equation, is negligibly small. At high pressures P plotted against $1/V$ should give a straight line from the slope of which β, the radius of molecular action, may be calculated. Kubo has applied the equation to results obtained by Uhlig, Kirkwood and Keyes[17] for several gases at pressures of the order of 100 atmospheres and temperatures of 100–200° C. They found that for H_2, N_2, and CH_4, which are non-polar, P was constant at constant temperature, but for NH_3, which is polar, there was a definite increase of P with pressure. The values given for three different temperatures were consistent with the above equation and gave a value for β which is in good agreement with values deduced from the viscosity measurements of Rankine and others, and from the van der Waals constant b. For CO_2 a slight increase of P with pressure had been recorded, and Kubo was not able to account for this by his theory since CO_2 is non-polar. However, measurements of high precision have now been made by A. Michels and others[18,19,20] on H_2, N_2, and CO_2 at pressures of 1000 atmospheres or more, and temperatures up to 100 or 150° C. The results show that there is little to be accounted for in non-polar gases. For each of these gases P was found to be independent of temperature

within the experimental error, and for H_2 and N_2 it was also independent of pressure. For CO_2 a tendency to decrease with increase of pressure is reported. The accuracy of the Clausius-Mosotti relation for these materials is quite striking, e.g. for N_2 the deviations from the constant value of P are less than would be caused by an error of 0·0001 in the value of ϵ for all pressures and temperatures giving values of ϵ up to 1·12. For the highest pressures the experimental errors are naturally greater, but the deviations are less than would correspond to an error of 0·001 in ϵ for values of ϵ up to 1·300. Valuable information concerning intermolecular forces may be expected from further work in this direction.

Experiments have also been made with the object of measuring any change of dielectric constant that may result from the application of a strong magnetic field to the material, but the results are almost entirely negative. A. C. Young[21], working with nitrogen and oxygen at a pressure of 100 atmospheres, and also in the liquid form, found no measurable effect of a field of 19,000 G. except an increase of ϵ for oxygen gas of about 1 in 10^8, which may be accounted for by an increase of pressure arising from the magnetic susceptibility. The sensitivity of the measurements on the liquefied gases was not so high on account of difficulties arising from temperature changes, but the effect was less than 1 in 10^5.

An interesting point arises from the work of Stranathan[11] on water vapour. Several previous observers had reported deviations from equation (5) when the vapour pressure was increased beyond a certain point. This effect has been variously attributed to association of the molecules of the vapour, adsorption of the molecules on the condenser plates, and errors in capacity measurement due to conductance in the test-condenser. Stranathan shows conclusively that the effect arises from the formation of an adsorbed film of water molecules on the solid insulating supports separating the two plate-systems of the test-condenser, the capacitance of this film (of dielectric constant 80) increasing the apparent value of ϵ. The errors may be minimized by the use of quartz insulators of considerable length. The capacitance of the surface film is then as small as possible. In this way the Debye law was verified for water vapour at pressures of 0–70 cm. Hg and temperatures of 20–200° C. Deviations only occurred near the saturation point.

In connexion with the dielectric constant of a film of water molecules adsorbed from vapour on the surface of a solid insulator, another interesting point arises. Argue and Maass[22] have made measurements of the increase of capacitance of a test condenser containing pure cellulose, initially quite dry, as the cellulose is allowed to take up successive known quantities of water. Although there must be some uncertainty as to the distribution of the cellulose fibres between the plates of the condenser, it is possible to estimate from the observations the dielectric constant of the cellulose itself, and of the water adsorbed at the various stages. The results show that the dielectric constant of the film first formed is of the order of 15, and that as the film increases in thickness the value of ϵ increases until it approaches the ordinary value for water, viz. 80. The explanation is to be found in the polar properties of the water molecule. The high dielectric constant of water in the ordinary liquid state is almost entirely due to the orientation of the polar

molecules. But the molecules that form the first adsorbed layer on the surfaces of the cellulose micelles are strongly held to the surface, so that their rotation is severely limited, with the result that the dielectric constant of the layer is low. The molecules in a second layer would not be as securely anchored, and would be free to rotate between wider limits. Thus the dielectric constant of successive layers increases, the molecules in the outermost layers having the same freedom of rotation as in ordinary water in bulk and therefore having the same value of ϵ. Stranathan estimated that the adsorbed layers on his quartz insulators were of the order of 100 molecules thick and that the dielectric constant of the layer could be considered as that of ordinary water. Observations consistent with those of Argue and Maass have been made by the writer[23] on varnished cloth (as used in the electrical industry for insulating purposes) which has been allowed to take up from the atmosphere various amounts of water. These experiments showed that the power factor of the material varies with moisture content in a similar way to the capacitance, which suggests that the conductivity of the inner adsorbed layers is also less than those of the outer layers.

Polar groups in relative rotation. In some cases Debye's equation is found not to hold, i.e. P is not proportional to $1/T$, but the material behaves as though the electric moment of its molecule varied with temperature. Examples are the ethylene halides. It is believed that such molecules contain two polar groups which are in relative rotation about a common axis, the C—C bond. When the potential energy between the two halves of the molecule is less than $0{\cdot}1kT$ this relative rotation is free and unhindered, and the moment of the molecule possesses a value which may readily be calculated from those of the two groups. If, however, the potential energy is greater than $0{\cdot}1kT$ there is a rotary oscillation about the position of minimum potential energy, which is a position of zero resultant moment. In this case the amplitude of oscillation will increase with rise of temperature, and therefore the mean molecular moment will increase with temperature, for the time average of the relative positions of the two groups will be farther and farther from the position of zero moment. Further information on this subject, including references, will be found in the survey by Williams[15].

Solutions of polar materials. We have seen that it is not unreasonable to expect that Debye's equation may be applied to dilute solutions of polar materials in a non-polar solvent, e.g. to nitrobenzene or water dissolved in, say, hexane or benzene, so that it should be possible to deduce the electric moment of the polar molecule from measurements of the dielectric constant of such solutions as well as from those of the vapour. Much experimental work has been done with solutions. As is to be expected, unless the solutions are very dilute their behaviour is complicated by association on the part of the polar molecules, and it therefore sometimes becomes necessary to work with solutions of varying small concentrations, and to obtain by extrapolation the properties corresponding to infinite dilution. When solutions are examined in this way it is usually found that the dielectric constant varies with temperature in accordance with Debye's equation and that it is possible to deduce values for the electric moment. Values obtained in this way agree fairly

well with those obtained from the vapour, but there is very frequently a small discrepancy which is greater than the experimental error. Several observers[24] have noted that the discrepancy appears to vary systematically with the dielectric constant of the solvent, and have proposed empirical formulae from which it may be calculated. The one most widely used is that of H. Müller[25]:

$$\mu_{\text{sol.}} = \mu_{\text{gas}} [1 - 0{\cdot}038\,(\epsilon - 1)^2].$$

This satisfactorily represents the observations for several materials in a variety of solvents, but it certainly does not apply to all substances; indeed the discrepancy is sometimes of the opposite sign.

Several possible explanations of these apparent differences of electric moment have been proposed. Debye[26] has interpreted them in terms of a quasi-crystalline structure of liquids. The local field acting on a molecule due to its neighbours is regarded as directed along an axis, which is independent of the applied field, and which changes slowly in comparison with the thermal agitation in liquids. This local field thus tends to prevent orientation of the polar molecules in the direction of the applied field, and this effect is added to that of thermal agitation. Thus the polarization due to orientation is smaller on account of the presence of the solvent, and the apparent dielectric constant is correspondingly smaller. This explanation may well apply to certain cases, but it only accounts for effects of one sign, and does not appear to connect them very closely with the dielectric constant of the solvent, which the experimental evidence shows to be the most important factor.

Raman and Krishnan[27] have also considered the possible anisotropy of liquids, and have derived a modified form of the Clausius-Mosotti relation in which the effects of anisotropy are represented by two additional constants. M. A. G. Rau[28] has applied this relation to the derivation of an expression for the molar polarization of an infinitely dilute solution of a polar material in a non-polar solvent, after eliminating the effect of the solvent, and shows that certain empirical formulae proposed for the effect of the solvent in such solutions may be derived as approximations from the expression obtained. Measurements on dilute solutions of nitrobenzene in various solvents made by Rau and Narayanaswamy[29] are also consistent with this expression.

A somewhat different line of attack on this problem was initiated by Weigle[30], who considered the effect of the electric field of a polar molecule on the non-polar molecules surrounding it. A simple calculation shows that near the surface of the polar molecule the field must be quite intense (of the order of 10^7 V./cm.) and one might perhaps expect the effects of such fields to be apparent in measurements made of the over-all properties of the solution. Weigle pointed out that if the surrounding non-polar molecules are anisotropic they will tend to set themselves so that their direction of easiest polarization lies in the direction of the local field. Such molecules will therefore be orientated by the dipole. An additional component of polarization is produced; in effect an additional electric moment is added to that of the dipole. This additional moment was estimated to be of the order of 1 per cent for the solutions usually examined and is negative for spherical molecules. If,

however, the polar molecules are not spherical, the polarization of neighbouring molecules may be equivalent to a much larger electric moment in the reverse sense. In one case this effect was calculated to amount to 15 per cent of the moment of the free molecule and it was considered to account for the observed effects of the various solvents used in measurements of electric moment. Higasi[31] has applied Weigle's theory to the results of measurements made with various solvents. He explains the more common negative effect as holding for molecules, which are longer in the direction of their dipole axis than in directions at right angles to this, while a positive effect is obtained for molecules which have their longest axis perpendicular to their dipole axis. It is immediately obvious that molecules which are end-on to the dipole will be polarized by its field in the opposite sense to those which are broadside-on, and that this difference introduces the possibility of an over-all effect of either sign. F. C. Frank[32] has continued this line of thought. He emphasizes the importance of the geometry of the polar molecule in determining the effect of any solvent. Thus a molecule may consist of a small polar group attached to a comparatively large radicle on the dipole axis, or alternatively the radicle may not be in the axis. Two such molecules would be polarized very differently in the vapour condition, but dissolved in a non-polar solvent the difference might be very much smaller since each solution would contain small polar groups completely surrounded by non-polar groups, the attached radicle in solution behaving in much the same way as a molecule of the solvent. Frank considers the nature and magnitude of the solvent effect for a number of molecules whose geometry may be considered to be known approximately. He treats the solvent as a continuous medium and obtains the total polarization by integration. He concludes that the solvent effect may be represented by

$$\Delta\mu/\mu = A\,(\epsilon - 1)/\epsilon,$$

where ϵ is the dielectric constant of the solvent and A is a factor which depends mainly on the geometry of the polar molecule. He calculates the effect for a few simple shapes and obtains results which account for much existing data. Müller's empirical rule is considered to apply to molecules with a radicle attached to their dipole axis, e.g. $C_6H_5.Cl$, $C_6H_5.NO_2$, etc. Results obtained with polar solvents are not explained by his calculations, but this could not be expected. Frank points out that the polarization of the solvent molecules by the dipoles produces several secondary effects, one of which is that the polar molecules attract the solvent molecules as a magnet attracts iron, with the result that a monomolecular layer of solvent probably becomes bound to the polar molecule. This may well be the explanation of the very great difficulty of removing the last traces of water from benzene.

Dispersion and absorption. The expressions for the dielectric constants of materials possessing polar molecules so far considered have been derived from a consideration of the statistical equilibrium resulting from the application of a constant electric force. It is obvious that a finite time will be required for the establishment of this equilibrium, and this may be represented by the time of relaxation τ, which may be considered as the time required for the fraction $1/e$ of

the molecules to resume a random distribution of orientation after the field is removed. In practice, measurements of dielectric constant are almost invariably made with alternating voltage, and it becomes obvious that the value of ϵ measured will only correspond to those of the above equations if the periodic time of the applied voltage is very large compared with the time of relaxation of the molecular system. When this is the case the polarization will be practically in phase with the applied voltage, and independent of the actual frequency so long as this is small compared with $1/\tau$. The value of ϵ so obtained may be called the static value ϵ_s. If now the frequency of the applied field is increased until it is very large compared with $1/\tau$, the molecules will no longer be able to reverse their mean orientation as the field is reversed: the reversals of field will occur so rapidly that the field is never established in one direction for a sufficient length of time to cause an appreciable departure from the random distribution of orientation. Thus the polarization due to molecular orientation becomes zero and the dielectric constant is correspondingly smaller. It is convenient to denote this value by ϵ_0, the value for very high, say optical, frequencies. We see, therefore, that the dielectric constant diminishes from the static value ϵ_s to the optical value ϵ_0 as the frequency of the applied field increases from a value small compared with $1/\tau$ to a value large compared with $1/\tau$, i.e. the material shows anomalous dispersion for electromagnetic radiation of frequencies of the order of $1/\tau$. At such frequencies it is clear that the mean molecular orientation will be appreciable, but less than the static value; moreover there will be a difference of phase between the electric field and the polarization due to orientation, which means that there will be a loss of energy in each cycle. The anomalous dispersion is therefore accompanied by absorption. Debye's expression[1] for the molar polarization at the frequency $\omega/2\pi$ is

$$\frac{\epsilon-1}{\epsilon+2}\frac{M}{P}=P(\omega)=\frac{4\pi N}{3}\left[\alpha_0+\frac{\mu^2}{3kT}\frac{1}{1+j\omega\tau}\right],$$

from which it follows that the complex dielectric coefficient $\epsilon'-j\epsilon''$ is given by

$$\epsilon'=\epsilon_0+\frac{(\epsilon_s-\epsilon_0)}{1+\beta^2\omega^2} \qquad \text{.....(6),}$$

$$\epsilon''=\frac{(\epsilon_s-\epsilon_0)\,\beta\omega}{1+\beta^2\omega^2} \qquad \text{.....(7),}$$

where $$\beta=\left(\frac{\epsilon_s+2}{\epsilon_0+2}\right)\tau.$$

The so-called real component may be regarded as the effective dielectric constant at this particular frequency. It is the value which corresponds to the equivalent shunt capacitance of a test-condenser containing the material. The so-called imaginary component ϵ'' is a measure of the energy dissipation. The loss tangent of the material at this frequency is therefore given by

$$\tan\delta=\frac{(\epsilon_s-\epsilon_0)\,\beta\omega}{\epsilon_s+\epsilon_0\beta^2\omega^2} \qquad \text{.....(8).}$$

Thus at the frequency determined by $\beta^2\omega^2=\frac{\epsilon_s}{\epsilon_0}$ the loss tangent rises to a maximum value given by

$$\tan\delta_{max.}=\frac{1}{2}\frac{\epsilon_s-\epsilon_0}{\sqrt{(\epsilon_s\epsilon_0)}}=\frac{(\epsilon_s+2)(\epsilon_0+2)}{27\sqrt{(\epsilon_s\epsilon_0)}}\frac{2\pi\mu^2 n}{kT},$$

where n is the number of polar molecules per cm^3.

The equivalent conductivity of the material is given by

$$\sigma=p\frac{(\epsilon_s-\epsilon_0)}{1+\beta^2\omega^2}\frac{\beta\omega^2}{},$$

where p is a constant depending on the system of units used. Thus the equivalent conductivity increases as the frequency rises, approaching the limiting value $p(\epsilon_s-\epsilon_0)/\beta$ at infinitely high frequencies.

The question now arises, what determines the time of relaxation τ of the system. Debye considered the simple system of a dilute solution of a polar material in a non-polar solvent. As a first approximation he regarded the polar molecules as spheres of diameter a. Their rotation will be opposed by an inner frictional force, akin to that of viscosity. Since the mobilities of ions in liquids have been found to be consistent with the notion that Stokes's law may be applied to bodies of that kind, it is not unreasonable to apply it tentatively also to polar molecules. In this way we obtain for the time of relaxation $\tau=4\pi\eta a^3/kT$, where η is the constant of internal friction. Assuming η to be the same as the ordinary coefficient of viscosity, Debye estimated that τ for water should be of the order of $0{\cdot}2\times10^{-10}$ sec., so that we might expect to observe the phenomena of dispersion and absorption at frequencies of the order of 10^{10} c./sec. Until very recently such frequencies were unknown in either electrical or optical practice, but Cleeton and N. H. Williams[33] have now observed absorption in ammonia at frequencies of this order. Their technique may be regarded as combined optical and electrical. Free electromagnetic waves were generated by means of an oscillator of the magnetron type, and directed on to an echelette diffraction grating by means of metal parabolic mirrors 3 ft. in diameter. The emergent beam was brought to a focus by means of a second similar mirror, and a crystal rectifier with sensitive galvanometer was used as detector. This stands out as an extremely interesting verification of Debye's theory applied to gases. However, it is scarcely to be expected that such measurements could be very accurate, and most investigators of this problem have acted upon a suggestion made by Debye, and have made measurements on dilute solutions of polar materials in non-polar solvents of high viscosity. The relaxation times of such systems may be expected to be very much larger than that of a mobile liquid such as water, and therefore dispersion and absorption are likely to be observable over a range of frequency for which the technique of accurate measurement is well established. Investigations of this type have recently been described by Luthi[34] and by W. Müller[35], both of whom have made measurements on nitrobenzene dissolved in various mineral oils of high viscosity; by Jackson[36], who used cetyl palmetate dissolved in paraffin wax, and by Sommerman[37], who used phenol and stearic acid

dissolved in a mixture of paraffins. All these observers obtained some evidence tending to support Debye's theory, although there was also evidence that the simple theory outlined above is not altogether adequate. For example, Luthi observed two regions of dispersion, at frequencies in the neighbourhood of 3×10^6 and 6×10^8 c./sec. respectively; both Jackson and Sommerman found that the peak of the curve connecting loss tangent and frequency was broader than equation (8) predicts, and although in certain instances the inner friction constant η appeared to vary with temperature in much the same way as does the ordinary coefficient of viscosity, yet W. Müller, Jackson and Sommerman all note that η must often be considered as 1000 times smaller than the macroscopic viscosity. Müller found the ratio of η to the ordinary viscosity to increase with temperature, the indication being that the two coefficients are identical at high temperatures. At the lower temperatures it should be stated that the solutions had solidified, and it is a remarkable fact that no pronounced change in the dielectric properties under consideration occurs during the change from the liquid to the solid state. The only change observed is an increase of relaxation time. The conception of polar molecules freely rotating in a solid solution is apt to seem somewhat extravagant at first, and one certainly does not expect the forces opposing rotation to be merely those of viscosity. Gemant[38] has therefore considered the application of Debye's theory to a solid which possesses elastic as well as viscous properties. The rotation of the molecules is considered to be controlled by an elastic controlling force, which decays exponentially with time. His conclusion is that Debye's equations are only slightly altered by the introduction of elasticity and that the application of the theory to such solids is quite justified. Yates-Fish[39] has also showed that Stokes's law may be applied to the motion of a body in such a medium merely by replacing the ordinary coefficient of viscosity by a coefficient which also depends on elasticity, the geometrical factors remaining unaltered.

Polar liquids and solids. We have seen that gases and vapours are most likely to satisfy the assumptions made in the development of Debye's theory, and that dilute solutions of polar materials in non-polar solvents may be expected to be second to gases in this respect. We might expect the motions of the molecules of pure polar liquids and solids to be much more complicated, but many observers have boldly applied Debye's equations to liquids and solids, and it is found that the behaviour of many materials may be represented by equations of the form given by Debye. It should be noted that this does not necessarily mean that the explanation provided by the theory is correct. The early work on liquids and solids has been reviewed by Debye, and a considerable amount of additional information has been obtained in the last year or two. Thus Morgan[40] has observed dispersion and absorption in glycerine at frequencies varying from 30 c./sec. to nearly 10^8 c./sec. The frequency of maximum absorption varied continuously over this range as the temperature varied, the maximum value of ϵ'' increasing slightly with decrease of temperature in accordance with the above equations. Jackson[41] has obtained very similar results with a chlorinated diphenyl, an insulating liquid, which has recently been produced commercially in large quantities. (It has many of the virtues of

ordinary transformer oil, but is non-inflammable and has a dielectric constant of about 5 as against 2·2 for hydrocarbon oils. Its general properties are described by Clark[42].) E. J. Murphy[43], confirming and extending earlier work of Errera[44], Wintsch[45] and Smyth and Hitchcock[46], has made measurements of the dispersion and absorption in ice at frequencies from 16 to 10^5 c./sec. The surprising fact is established that the behaviour of ice at temperatures down to say $-100°$ C. is very similar to what one would expect on the above theory of a viscous liquid containing polar molecules. The static value ϵ_s of the dielectric constant of ice is not very different from that of water. It increases as the temperature falls, reaching a value of 145 at $-90°$ C., and then falling gradually to 2 at $-170°$ C. This last value is also the optical value ϵ_0 at all temperatures, and we may infer that at this very low temperature the polar molecules are no longer free to rotate. Dispersion, i.e. the change of ϵ with frequency from ϵ_s to ϵ_0, was found by Murphy to occur at frequencies of the order of 100 c./sec. at $-46°$ C. and so on, just as if the relaxation time were controlled by a viscosity which diminished rapidly with rise of temperature. The curves of both ϵ' and ϵ'' were found to be consistent with Debye's equations (6) and (7), the time of relaxation increasing exponentially with diminishing temperature, from 2×10^{-5} sec. at $-0·8°$ C. to about 20 sec. at $-139°$ C. There was, however, evidence of two relaxation times having values, e.g. of 2 sec. and 100 sec. at $-116°$ C., so that the interpretation of the results is not quite straightforward, even if one is prepared to accept the notion of an internal viscosity for ice. In connexion with the difficulty of imagining any rotation on the part of molecules in a crystal lattice, Debye has pointed out that if the moment of the H_2O molecule in ice and water is the same as that in water vapour, then the static dielectric constant of $\epsilon_s = 80$ would be accounted for if an applied field of 1 V./cm. caused the complete rotation (reversal of orientation) of only 1 molecule in each five million. Thus the actual range of rotation of any one molecule need only be extremely small in order to account for the observed facts. Sun and Williams[47] have examined several other materials in the form of crystalline solids with the idea of detecting any similar dispersion that might occur in them. Their range of frequency was 3×10^2 to 6×10^5 c./sec., and in this range no dispersion was detected in solid benzene and ammonia; there was a little in nitrobenzene, and it was more pronounced in benzophenone and acetophenone, but not nearly so marked as in water. The changes of ϵ' with frequency for these substances satisfied a formula of the type (6), but the changes were small and their interpretation uncertain as several possible mechanisms of dispersion and absorption give the same type of formula, and there is not yet sufficient evidence to enable us to discriminate between them.

Other theories of dispersion and absorption. It is perhaps not immediately obvious that there is no essential difference between the absorption and accompanying dispersion discussed in the previous paragraphs, and the dielectric absorption associated with the phenomena of the residual charge and the anomalous charge and discharge currents of condensers. The differences are merely those of time or frequency, and as we have seen that the time of relaxation in one and the same substance may vary continuously from several seconds to 10^{-8} sec. say, there seems

no great justification for regarding the long-time phenomenon of dielectric absorption currents as different from the short-time phenomenon of the absorption of electromagnetic radiation. Thus any of the older theories of residual charge and dielectric absorption will account for a change of apparent dielectric constant with frequency, i.e. for dispersion, and the formulae expressing ϵ' and $\tan\delta$ as functions of frequencies are usually of the same form as those given by Debye's theory. Indeed these formulae are common to all theories which assume that the motion of the carriers of electricity which produces the polarization is opposed by a restoring force proportional to the displacement and also by a frictional force proportional to the velocity. A summary[48] of the older theories up to 1926 has been given by the writer, who has also described an additional possible mechanism[49], viz. the accumulation of ions of opposite sign near the surfaces of the two electrodes. The back e.m.f. produced by these charges provides the necessary restoring force. More recently, Murphy and Lowry[50] have offered another theory for solids, while Debye's theory of inter-ionic attraction in electrolytes provides yet another theory for fluids. The picture given by Murphy and Lowry for crystalline solids is as follows. The substance is built up of a number of minute structural units, each containing, say, 10^4 molecules. Each unit is a perfect insulator; dielectric imperfections are confined to the spaces between the units, and any water which may be absorbed is accumulated there. These spaces may be regarded as conducting closed surfaces insulated from one another, and the movement of adsorbed ions on the surfaces is the cause of absorption and dispersion. Such adsorbed ions cannot leave the surfaces to which they are attached, but under the influence of an applied field, the positive ions move over the surface in one direction and negative ions in the opposite direction, thus giving each structural unit an electric moment. It is clear that the properties of such a system will be in many respects similar to those which would obtain if each structural unit possessed a permanent electric moment and were capable of a change of orientation under the influence of an applied field, and it is to be expected that the equations for dispersion and absorption will be of the same form as those for Debye's theory. The chief difference lies in the fact that the relaxation time will be determined by the mobilities of the ions over the surface, instead of the "inner viscosity" of the material.

Debye's theory of inter-ionic attraction for electrolytes is based on the electrostatic attraction between ions of opposite sign, as a result of which each ion tends to surround itself with those of opposite sign, and there is therefore a certain regularity of distribution of positive and negative ions. In the development of the theory this regularity is represented by considering each ion to be surrounded by an atmosphere of ions of opposite sign, and it is shown that these ionic atmospheres must be regarded as possessing a certain characteristic time of relaxation, which represents the natural time of formation or disintegration of the atmosphere. As a result of this time of relaxation the motion of an ion under the influence of an applied field is opposed by an additional component of force besides that due to viscosity; for the atmosphere of an ion which is not drifting under the action of an external field may be regarded as possessing spherical symmetry, so that the force

between the ion and its atmosphere has the same value in every direction, but when the ion begins to drift under the action of a field, this symmetry is destroyed owing to the finite time required for the atmosphere to assemble and disintegrate at the points successively occupied by the ion in its motion. It is almost as though there were a slight lag between the ion and its atmosphere (though the motion of the atmosphere must be thought of as like that of a wave and not as one of translation of the ions forming it). The result is that the force exerted by the atmosphere on the ion is no longer the same in all directions: there is a resultant force in the opposite direction to the motion of the ion. This force is often called the relaxation force. When the applied field is alternating it is obvious that the relaxation force will vary with the frequency, for when the frequency is very high compared with $1/\tau$, the atmosphere will be no longer able to follow the motion of the ion. Effectively, we then have a stationary atmosphere with the ion vibrating at its centre, the mean force exerted by the atmosphere on the ion is now the same in all directions, i.e. the relaxation force has vanished. At frequencies very low compared with $1/\tau$ the relaxation force has its full value. Thus, as the frequency is increased through values of the order of $1/\tau$, the conductivity increases, approaching the limiting value corresponding to zero relaxation force. It will be remembered that this increase of conductivity is also a characteristic of absorption. The mathematical theory of the above process which has been worked out by Debye and Falkenhagen[51] shows that the dielectric constant of the electrolytic solutions diminishes with rise of frequency as the conductivity increases, i.e. the absorption is accompanied by dispersion.

A consideration of the foregoing theories shows clearly that the existence of absorption and dispersion in a dielectric is no proof of the occurrence of dipole action. Several other possible mechanisms may produce the same results, and in those cases in which two relaxation times were observed in solids it is quite likely that two different mechanisms were involved.

Experiments on electrolytic solutions. Several observers have recently made measurements of the dielectric constants of electrolytic solutions and compared the values with those of the pure solvent. Generally speaking the results have been in accordance with the Debye-Falkenhagen theory mentioned above. The dielectric constant of the solution was greater than that of the pure solvent and the difference increased with the concentration and with the valency for a given molar concentration, but diminished as the frequency was increased to very high values. The high conductivity of electrolytic solutions makes the accurate measurement of their dielectric constant somewhat difficult, but it is interesting to note that the above results have now been confirmed by several observers using different methods. E. Fischer[52] used a bridge method at frequencies of 10^8–10^9 c./sec. Resonance methods with detectors of various kinds were used by R. W. McKay[53], M. Röver[54] and Williams and Arnold[55]. W. Hackel[56] used a barretter method, while A. Dember[57] employed stationary waves produced between a parabolic reflector and a plane mirror, and determined the index of refraction by measurements of wave-length. It is curious that when aqueous solutions of electrolytes were first

measured, it was expected that the dielectric constant of the solution might be lower than that of the pure solvent, owing to the polar character of the latter, for the polarization due to dipole orientation cannot exceed the limiting value in which all the dipoles are in line with the field, and as the applied field is increased to very high values the approach to this saturation value will be made evident by a diminution of dielectric constant. The fields obtainable by ordinary laboratory methods are far too small to produce such a saturation effect in water, but the field due to an ion is very intense in its immediate vicinity and it was expected that the saturation effect would be appreciable in the water immediately surrounding the ion, and that therefore the dielectric constant of the solution would be somewhat smaller than that of the solvent containing no ions. Early workers appeared to have obtained some evidence of this saturation effect, but all the observers mentioned above agree that the effect cannot be detected.

The dielectric analogue of ferromagnetism: Rochelle salt. The theory of polar molecules so far considered is in the main restricted to cases in which the Clausius-Mosotti relation is valid, and we have seen that there is little justification in applying this to solids and liquids in which the internal field may be very large and not at all likely to satisfy the relation. Two questions suggest themselves. To what extent does the large internal field affect the freedom of rotation of the dipoles? and do the electrical dipoles in solids ever behave in a manner analogous to that of the magnetic dipoles in iron? Pauling[58] provided a partial answer to the first question. He maintains that the motion of polar molecules may approach two limiting cases. When the intermolecular forces are small we have the free rotation already considered. When these forces are large, the rotation is severely hindered and we have only oscillation about equilibrium positions. A solid may exist in two states corresponding to these two cases, and a transition from one state to the other, often accompanied by a change of crystal-structure, occurs over a definite range of temperature. Changes of specific heat occur in such a region and from such data it was predicted that solid HCl would, at temperatures just below the melting-point, have a high dielectric constant which would vary greatly with temperature (molecular rotation), while at lower temperatures the value would be much smaller and nearly independent of temperature (molecular oscillation). This was verified by several experimenters. The behaviour of ice mentioned previously is very similar.

The question as to the possibility of the existence of dielectric properties analogous to those of ferromagnetism has been answered by several investigations of Rochelle salt. The dielectric constant of this material in the direction of the *a* axis may reach values as high as 10^4. The value is variable with the strength of the applied field. The polarization increases linearly with increasing field strength up to 50 V./cm. corresponding to $\epsilon = 400$. From this point it increases very rapidly to a value corresponding to $\epsilon = 11{,}000$ at 100 V./cm., and then at 300 V./cm. reaches saturation. If the field is now reversed, a hysteresis loop may be traced. These properties are closely analogous to those of soft iron, but there is a difference in that hysteresis only occurs between the critical temperatures -20 and $25°$ C.

A detailed summary of the properties of Rochelle salt is given by Staub[59]. A model by which some of these properties may be explained has been suggested by R. H. Fowler[60]. The polar molecules in Rochelle salt $NaKC_4H_4O_6.4H_2O$ are those of the water of crystallization. When these molecules are capable of orientating freely under the influence of the applied field, we have the very high values of ϵ. The upper critical temperature corresponds to the Curie point, and is explained by the Weiss-Langevin theory of ferromagnetism. The existence of the lower critical temperature must mean that below this point the dipoles are no longer capable of free rotation, i.e. the orientations are no longer free but restricted by the local field of neighbouring molecules. Fowler points out that the effect of the rotations of these molecules must be to diminish this local field, i.e. the local field depends on the amount of rotation already present and vanishes when this is sufficiently large. Following this idea, a model is suggested in which the molecular dipoles are arranged in strings parallel to an axis of the crystal, and it is shown that this model accounts for the existence of a very high dielectric constant within a range bounded by upper and lower critical temperature and also for the anisotropy of the crystal. It will be very interesting to note in due course whether X-ray analysis confirms the distribution of water molecules in this manner in the crystal. Fowler's theory is not concerned with piezo-electric effects, but Müller[61] and others have shown that the dielectric properties are considerably modified if piezo-electric deformations are prevented, and a complete theory of Rochelle salt must take this into account. Müller[61] has made measurements of several kinds, dielectric and optical, on Rochelle salt, and has shown that nearly all the existing information may be correlated by a phenomenological theory based on the internal field $F=E+fP$, the polarization P containing four components, (1) displacement of electrons in the molecule, (2) displacement of ions in the lattice, (3) a back polarization due to the piezo-electric effect, and (4) dipole action. Component (3) vanishes if deformations are prevented. The dipole action is considered to be analogous to ferromagnetism with the difference that the Curie point is regarded as variable with temperature, i.e. the dielectric properties at the temperature T are analogous to the magnetic properties of a ferromagnetic material whose Curie point is a function of T. This simple assumption provides a remarkable correlation of the varied experimental data. Another aspect of the properties of Rochelle salt has been investigated by Schulwas-Sorokin and Posnov[62]. They have considered time effects and frequency effects. The polarization due to a constant applied mechanical stress is shown to be not instantaneous, but to possess a time of relaxation of the order of 10 sec. The polarization for an applied electric field has a different time of relaxation of the order of 0·01 sec. The dielectric constant and power factor of the material was found to vary greatly with frequency and the main features of the result could be accounted for by the two times of relaxation. There were, however, considerable unexplained variations.

It is clear that the properties of Rochelle salt offer a most fruitful field to the student of dielectrics, and there will no doubt be further developments in this direction.

References

(1) DEBYE, P. *Polar Molecule.* Chemical Catalog Co. New York (1929).
(2) *Trans. Faraday Soc.* **30**, Appendix (1934).
(3) KUBO, M. *Sci. Pap. Inst. phys. chem. Res.*, Tokyo, **26**, 242 (1935).
(4) KUBO, M. *Sci. Pap. Inst. phys. chem. Res.*, Tokyo, **27**, 65 (1935).
(5) MAHANTI, P. C. *Z. Phys.*, Lpz., **94**, 220 (1935).
(6) SMYTH, C. P. and MCALPINE, K. B. *J. chem. Phys.* **3**, 347 (1935).
(7) MCALPINE, K. B. and SMYTH, C. P. *J. chem. Phys.* **3**, 55 (1935).
(8) MAHANTI, P. C. *Phil. Mag.* **20**, 274 (1935).
(9) GROVES, L. G. and SUGDEN, S. *J. chem. Soc.* p. 971 (1935).
(10) SÄNGER, R. *Phys. Z.* **31**, 306 (1930).
(11) STRANATHAN, J. D. *Phys. Rev.* **48**, 538 (1935).
(12) MÜLLER, F. H. *Phys. Z.* **35**, 1009 (1934).
(13) ABADIE, P. and CHAMPETIER, G. *C.R. Acad. Sci.*, Paris, **200**, 1590 (1935).
(14) DE BRUYNE, J. M. A. and SMYTH, C. P. *J. Amer. chem. Soc.* **57**, 1203 (1935).
(15) WILLIAMS, J. W. *J. Franklin Inst.* **219**, 47 (1935).
(16) KUBO, M. *Sci. Pap. Inst. phys. chem. Res.*, Tokyo, **27**, 295 (1935).
(17) UHLIG, H. H., KIRKWOOD, J. G. and KEYES, F. G. *J. chem. Phys.* **1**, 155 (1933).
(18) MICHELS, A., SANDERS, P. and SCHIPPER, A. *Physica*, 's Grav., **2**, 753 (1935).
(19) MICHELS, A., SANDERS, P. and SCHIPPER, A. *Physica*, 's Grav., **1**, 627 (1934).
(20) MICHELS, A. and MICHELS, C. *Philos. Trans.* **231**, 409 (1933).
(21) YOUNG, A. C. *Canad. J. Res.* **13**, 111 (1935).
(22) ARGUE, G. H. and MAASS, O. *Canad. J. Res.* **13**, 156 (1935).
(23) HARTSHORN, L. *J. Instn elect. Engrs*, **70**, 417 (1932).
(24) *Trans. Faraday Soc.* **30**, 707, 719, 729, 745 and 751 (1934).
(25) MÜLLER, H. *Phys. Z.* **34**, 689 (1933).
(26) DEBYE, P. *Phys. Z.* **36**, 100 and 193 (1935).
(27) RAMAN, C. V. and KRISHNAN, K. S. *Proc. roy. Soc.* A, **117**, 589 (1928).
(28) RAU, M. A. G. *Proc. Ind. Acad. Sci.* A, **1**, 498 (1935).
(29) RAU, M. A. G. and NARAYANASWAMY, B. N. *Proc. Ind. Acad. Sci.* A, **1**, 489 (1935).
(30) WEIGLE, J. *Helv. phys. Acta*, **6**, 68 (1933).
(31) HIGASI, K. *Sci. Pap. Inst. phys. chem. Res.*, Tokyo, **25**, Nos. 525–8 (1934).
(32) FRANK, F. C. *Proc. roy. Soc.* A, **152**, 171 (1935).
(33) CLEETON, C. E. and WILLIAMS, N. H. *Phys. Rev.* **45**, 234 (1934).
(34) LUTHI, R. *Helv. phys. Acta*, **6**, 139 (1933).
(35) MÜLLER, W. *Ann. Phys.*, Lpz., **24**, 99 (1935).
(36) JACKSON, W. *Proc. roy. Soc.* A, **150**, 197 (1935).
(37) SOMMERMAN, G. M. L. *J. Franklin Inst.* **219**, 433 (1935).
(38) GEMANT, A. *Phil. Mag.* **19**, 746 (1935).
(39) YATES-FISH, N. L. *Phil. Mag.* **21**, 226 (1936).
(40) MORGAN, S. O. *Trans. Amer. elektrochem. Soc.* **65**, 109 (1934).
(41) JACKSON, W. *Proc. roy. Soc.* A, **153**, 158 (1935).
(42) CLARK, F. M. *Trans. Amer. elektrochem. Soc.* **65**, 59 (1934).
(43) MURPHY, E. J. *Trans. Amer. electrochem. Soc.* **65**, 133 (1934).
(44) ERRERA, J. *J. Phys. Radium*, **5**, 304 (1924).
(45) WINTSCH, H. *Helv. phys. Acta*, **5**, 126 (1932).
(46) SMYTH, C. P. and HITCHCOCK, C. S. *J. Amer. chem. Soc.* **54**, 4631 (1932).
(47) SUN, C. E. and WILLIAMS, J. W. *Trans. Amer. chem. Soc.* **65**, 121 (1934).
(48) HARTSHORN, L. *J. Instn elect. Engrs*, **64**, 1152 (1926).
(49) HARTSHORN, L. *Proc. phys. Soc.* **37**, 215 (1925).
(50) MURPHY, E. J. and LOWRY, H. K. *J. phys. Chem.* **34**, 598 (1930).
(51) DEBYE, P. and FALKENHAGEN, H. *Phys. Z.* **29**, 401 (1928).
(52) FISCHER, E. *Phys. Z.* **36**, 585 (1935).
(53) MCKAY, R. W. *Canad. J. Res.* **12**, 377 (1935).

(54) Röver, M. *Ann. Phys.*, Lpz., **21**, 320 (1934).
(55) Williams, J. W. and Arnold, O. M. *Acta Physicochimica, U.R.S.S.* **3**, 619 (1935).
(56) Hackel, W. *Phys. Z.* **36**, 220 (1935).
(57) Dember, A. *Ann. Phys.*, Lpz., **23**, 507 (1935).
(58) Pauling, L. *Phys. Rev.* **36**, 430 (1930).
(59) Staub, H. *Naturwissenschaften*, **23**, 728 (1935).
(60) Fowler, R. H. *Proc. roy. Soc.* **149**, 1 (1935).
(61) Müller, H. *Phys. Rev.* **47**, 175 (1935).
(62) Schulwas-Sorokin, R. D. and Posnov, M. V. *Phys. Rev.* **47**, 166 (1935).

§ 2. ALLOYS FOR PERMANENT MAGNETS

By D. A. Oliver, B.Sc.

The older alloys. Most physicists are familiar with the well-known 6 per cent tungsten steel which is usually attributed to Lord Kelvin, who employed it for the needles of his famous marine compasses. Later this was studied exhaustively by Eversheds[63, 64], who was one of the first to carry out a comprehensive research on any one of the permanent magnet materials and who, with Watson[65], laid the foundations of all rational permanent magnet design.

The carbon steels containing chromium also appeared, but the next advance of major importance was made by Honda and Saito[66], who investigated the addition of cobalt to steels containing chromium and evolved very successfully a series of cobalt and cobalt-chromium steels culminating in the now well-established 35 per cent cobalt steel widely used for permanent magnets in instruments. All these steels require special heat-treatments, some of which are very much more complicated than others, but an underlying requirement of all is that the resulting crystal lattice shall be in a state of distortion or mechanical strain set up by the formation of carbides in a preferred state. This has been fully verified by Wood[67, 68] in the case of tungsten magnet steel. It would appear that in the 35 per cent cobalt steel this type of carbide-strain mechanism has been exploited to the fullest possible extent and much experimentation has not resulted in any marked advance being made by the addition of other elements. There is thus a clear line of demarcation between these permanent magnet steels having carbide structures and the carbon-free alloys now about to be described, which were briefly mentioned in volume 1 of these *Reports* (1934).

The newer alloys. In 1931 a discovery by Prof. Mishima[69] led to the development of an entirely new type of permanent magnet material. He was experimenting with ternary alloys of nickel, iron and a third element, and when this element was aluminium, for certain compositions he found that his test pieces showed abnormal mechanical hardness and brittleness. His investigations included an approximate examination of magnetic properties, and it was noticed that these alloys possessed quite exceptionally large hysteresis loops and above all phenomenally large values of coercive force as judged by the standards of existing materials.

Alloys exhibiting useful magnetic properties from a permanent magnet standpoint were found to lie within the following ranges: nickel 17–34 per cent, aluminium 8–14 per cent, iron the remainder. The highest values of stored magnetic

energy have later been found to be associated with a composition of the order of nickel 29 per cent, aluminium 12·5 per cent, iron the remainder, and typical magnetic values are remanence 5500 G., coercive force 500 oersteds, $(BH)_{max.}$ $1{\cdot}2 \times 10^6$ (average). (The maximum stored energy per cm³ is $(BH)_{max.}/8\pi$ erg.) This alloy can therefore be rightly associated with Mishima, although his earlier compositions contained less aluminium, but in commercial production the above composition is modified in that between 3 and 6 per cent of the nickel is replaced by copper, which ensures greater uniformity in the ultimate magnetic properties by making the heat-treatment less critical. These alloys pass severally under the names of "Alni" and "Nial", or simply "Nickel-aluminium". A further modification employs in its composition approximately 32 per cent nickel with no copper and when cooled more rapidly in heat-treatment than "Alni" can yield slightly higher values of $(BH)_{max.}$ and coercive force but much lower values of remanence. A smaller ratio of length to cross-section is therefore demanded in the design of the magnet part, which in certain circumstances is an advantage.

Further important advances were made in Sheffield by Horsburgh and Tetley[70] in 1934, when they studied the effects of the addition of cobalt to the nickel-aluminium-iron system. Mishima had previously mentioned cobalt as an improving element, but added alone the results were disappointing. Horsburgh and Tetley, however, overcame these difficulties and found that the addition of copper was essential. The composition for optimum magnetic results is critical and much work had to be carried out before it was ascertained. A typical analysis of this new alloy, now usually called "Alnico", is nickel 18 per cent, aluminium 10 per cent, cobalt 12 per cent, copper 6 per cent, iron 54 per cent, and typical average magnetic constants are: remanence 7500 G., coercive force 500 oersteds, $(BH)_{max.}$ $1{\cdot}6 \times 10^6$ (average). This alloy possesses the highest commercially obtainable energy content and is now firmly established as the best permanent magnet alloy in this class of available materials. In certain circumstances where space is not critical, the earlier alloys are slightly more economic, but the use of "Alnico" is increasing on account of its higher remanence, and energy content per unit volume. It is also superior mechanically as it is not so brittle as "Alni" and for certain special requirements can be turned and drilled after suitable rather lengthy heat-treatment.

Other new alloys. All the above alloys fall into the general category of dispersion-hardening alloys in which one phase is distributed throughout another phase on a diffuse submicroscopic scale. The theoretical treatment for this class of alloy is a quench from a high temperature followed by a tempering process, which usually greatly increases the mechanical hardness. In practice many of these alloys are too brittle to permit of a quench and recourse is had to a hybrid treatment.

About the same time as Mishima's discovery, Rogers and Seljesater[71] in America and Köster[74] in Germany found that carbon-free alloys could have outstanding properties from a permanent magnet standpoint. Exhaustive studies by Köster and Tonn[72,73] and Köster[74,75,76,77] on ternary systems throw much light on the structure of these newer permanent magnet alloys. A carbon-free ternary alloy of iron and cobalt with either tungsten or molybdenum can possess quite appreciable

magnetic properties, a typical composition being iron 67 per cent, cobalt 15 per cent, and tungsten 18 per cent. When this alloy is quenched from 1300° C. and tempered at 700° C. it can have a coercive force of 300 oersteds, a remanence of 10,000 G. and a value of $(BH)_{max.}$ approximately equal to $1{\cdot}3 \times 10^6$. These properties rival those of the ternary Mishima alloy, but the high cost of the elements employed precludes a widespread use of this particular alloy.

Very good magnetic properties have been secured by Honda, Masumoto and Shirakawa[78] from a ferrous alloy of cobalt, nickel and titanium. The ranges of composition are nickel 10–25 per cent, cobalt 15–36 per cent, titanium 8–25 per cent, iron the remainder, and coercive forces of 800–900 oersteds and remanence values of 7500–6400 G. are claimed within the above specified ranges. The values of $(BH)_{max.}$ for some of the experimental test pieces reached the figure of $2{\cdot}0 \times 10^6$, which therefore slightly exceeds the value of $1{\cdot}6 \times 10^6$ for alnico. The energy per unit of cost is lower and therefore the alloy is primarily important for its intrinsic scientific interest.

The highly interesting general class of dispersion-hardening alloys is being actively subjected to X-ray examination. Several papers[79, 80, 81] have already appeared, but the structures are so complex that it would be premature to comment on them at this stage, and much further work remains to be carried out before any really reliable correlations can be made. Bradley and Rodgers[82], however, using the X-ray method, have made a noteworthy contribution to the elucidation of the crystal structure of the Heusler alloys.

Powdered-alloy magnets. A different line of research was followed by Kato and Takei[83], who mixed the oxides of iron and ferro-cobalt (Fe_3O_4 and $Co.Fe_2O_4$), and compressed the mixture into the desired form of magnet, which was then heated to 1000° C. and magnetized at 300° C. to secure the best results. The remanence values are poor (3000–5000 G.), but coercive forces of 400–600 oersteds are claimed. The magnets are brittle but the density is low, about 4. An estimated value of $(BH)_{max}$ is $0{\cdot}9 \times 10^6$ and this is seen to be inferior to the ternary Mishima alloy.

A recent further development is the production in dies of magnets of compressed powdered alloys (such as alnico) mixed with a binder (such as bakelite powder). A final low-temperature treatment causes the binder to fuse and to be effective. This method overcomes certain casting and manipulative difficulties but is bound to result in a loss of effective remanence and in a lowering of the stored magnetic energy content per unit volume.

Finally it is not proposed to go into the design detail of actual magnet arrangements because that would carry us into the technology of the subject; suffice it to say that galvanometers, string oscillographs, indicating measuring instruments (such as voltmeters and ammeters), and the damping magnets of electricity energy meters, are all with advantage having their magnet systems redesigned in these new alloys. In the majority of instances the flux is increased, in other cases the same performance is retained with a smaller and lighter permanent magnet system.

References

(63) EVERSHED, S. *J. Instn elect. Engrs*, **58**, 780 (1920).
(64) EVERSHED, S. *J. Instn elect. Engrs*, **63**, 725 (1925).
(65) WATSON, E. A. *J. Instn elect. Engrs*, **61**, 641 (1923).
(66) HONDA, K. and SAITO, S. *Sci. Rep. Tohoku Univ.* **5**, 417 (1920); also *Phys. Rev.* **16**, 494 (1920).
(67) WOOD, W. A. *Phil. Mag.* **13**, 355 (1932).
(68) WOOD, W. A. and WAINWRIGHT, C. *Phil. Mag.* **14**, 191 (1932).
(69) MISHIMA, T. *Ohm*, **19**, No. 7 (1932).
(70) HORSBURGH, G. D. L. and TETLEY, F. W. Brit. Pat. Nos. 431,660, 439,543.
(71) ROGERS, B. A. and SELJESATER, K. S. *Trans. Amer. Soc. Steel Treat.* **19**, 553 (1931).
(72) KÖSTER, W. and TONN, W. *Arch. Eisenhüttenw.* **5**, 431 (1932).
(73) KÖSTER, W. and TONN, W. *Arch. Eisenhüttenw.* **5**, 627 (1932).
(74) KÖSTER, W. *Arch. Eisenhüttenw.* **6**, 17 (1932).
(75) KÖSTER, W. *Z. Elektrochem.* **38**, 549 (1932).
(76) KÖSTER, W. *Stahl u. Eisen*, **53**, 849 (1933).
(77) KÖSTER, W. *Arch. Eisenhüttenw.* **7**, 257 (1933).
(78) HONDA, K., MASUMOTO, H. and SHIRAKAWA, Y. *Sci. Rep. Tohoku Univ.* **23**, 365 (1934).
(79) GLOCKER, R., PFISTER, H. and WEST, P. *Arch. Eisenhüttenw.* **8**, 561 (1935).
(80) BURGERS, W. G. and SNOEK, J. L. *Physica*, 's Grav., **2**, 1064 (1935).
(81) JELLINGHAUS, W. *Z. tech. Phys.* **17**, 33 (1936).
(82) BRADLEY, A. J. and RODGERS, J. W. *Proc. roy. Soc.* A, **141**, 340 (1934).
(83) KATO, Y. and TAKEI, T. *J. Inst. elect. Engrs*, Japan, **53**, 408 (1933).

§ 3. SOFT MAGNETIC MATERIALS AND MAGNETIC MEASUREMENTS

By C. E. WEBB, B.Sc., A.C.G.I., D.I.C., A.M.I.E.E.

Soft magnetic materials. Since the principal factor in the production of great magnetic hardness is the existence of intense internal strains, it follows that in the development of materials having small coercive force and high permeability, the first consideration is the avoidance of such strains. This involves the elimination of impurities liable to cause lattice distortion, the most important of which are carbon and oxygen. Yensen and Ziegler[84] have determined the relations between the permeability of iron and the percentage contents of these elements down to 0·001 per cent, and have obtained very pure specimens with maximum permeabilities ranging up to 220,000.

Another important cause of internal strains affecting the permeability of a material is the magnetostrictive change in dimensions which occurs on magnetization. The importance of this factor was first pointed out by McKeehan[85] in connexion with the occurrence of the highest permeability in the permalloy series at approximately the composition for which the magnetostriction is zero. Von Auwers and Neumann[86], in an extensive survey of the effects of two different heat treatments on iron-nickel-copper alloys, have demonstrated a similar relationship in the ternary system. They showed the existence of a series of compositions, varying with the heat treatment, for which the initial permeability is exceptionally

high (up to 13,500), and a second series closely related to, but not identical with, the first, for which the coercivity is very small (down to 0·02 oersted), both series approximating to the range of compositions for which the magnetostriction is zero. An even more striking illustration of the influence of magnetostrictive strains on permeability is afforded by the large modifications produced in the properties of iron-nickel and iron-nickel-cobalt alloys by cooling them from a high temperature in a magnetic field, which have been described by Kelsall[87], and by Dillinger and Bozorth[88]. The maximum effect is obtained in 65-permalloy (containing 65 per cent nickel), which when cooled from 650° C. in a magnetic field of 10 oersteds increased in maximum permeability from 20,000 to 610,000—the highest permeability ever recorded—while its coercivity fell to 0·012 oersted. It must be noted, however, that the material in this state is very anisotropic, the increased permeability being obtained only parallel to the field applied during cooling and the permeability in a perpendicular direction being diminished. The effect is attributed to the relief of magnetostrictive stresses by plastic flow at high temperature in the individual Weiss domains. Subsequent re-magnetization in the same direction is thus facilitated while transverse magnetization is rendered more difficult.

An important practical improvement in magnetic materials is the production by Goss[89] of silicon-iron sheet having properties closely resembling those of single iron crystals, viz. high maximum permeability (more than 20,000) and a high permeability in strong fields in the direction of rolling, together with very low permeability in the perpendicular direction. Examinations of this material by Sixtus[90] and Bozorth[91] have shown that the magnetic anisotropy is due to strong preferred orientation of the crystals in sheets prepared by Goss's process, such that the direction of rolling approximately coincides with a direction of easy magnetization in the majority of the crystals.

Measurements on ferromagnetic materials. (*a*) *Double-yoke magnet-steel tester.* The development of permanent magnet materials with much higher coercive forces than were previously obtainable, described in § 2, has led to the use of such materials in the form of short, thick bars. The combination of high coercivity and small ratio of length to transverse dimensions has rendered many of the standard forms of testing apparatus unsuitable for use with these materials. A tester designed to meet these conditions, which, while embodying no fundamentally new principle, provides a novel combination of well-known principles of magnetic measurement, has been described by Stäblein and Steinitz[92]. Two E-shaped soft iron yokes are placed opposite one another with an air-gap between each of the three pairs of transverse limbs. Equal magnetizing windings are placed on each half of the long arms of both yokes, the windings being so connected that their m.m.fs. are in the same sense round the circuit formed by the outer limbs. Owing to the symmetry of the arrangement no flux crosses the centre air-gap in the absence of a specimen. If the symmetry is upset, however, by the introduction of a specimen into one of the two outer gaps, an additional flux is set up which completes its circuit mainly across the central air-gap, the flux in which is thus closely proportional to the value of $(B-H)$ in the specimen. By means of a suspended moving-

coil, through which an adjustable current flows, placed in the central gap, a deflection proportional to the gap flux, and hence approximately to $(B-H)$ for a given specimen, is obtained. Calculation shows that the relation between $(B-H)$ and the deflection, δ, is of the form

$$(B-H)=\delta\frac{C_1(1+C_2/l)}{i.A},$$

where l, A are the length and cross-section of the specimen, i is the current in the moving-coil, and C_1, C_2 are constants which may be determined by calibrating with two specimens of the same material having known magnetic properties but of different lengths. It follows that, if i is always adjusted proportionally to $C_1(1+C_2/l)/A$, a constant calibration for all specimens may be obtained.

The strength of the field applied to the specimen is measured by means of a small search-coil placed close to the specimen and rotated by a synchronous motor. The alternating e.m.f. generated is rectified by means of a commutator and applied to a d.c. millivoltmeter, the reading of which, when the position of the commutator brushes is adjusted to give complete rectification, is proportional to the field-strength, H. The whole apparatus for measuring H may be calibrated in any known field, and is found to give accurate readings of the value of H applied to the specimen, provided care is taken to avoid disturbance of the field through the presence of iron in the accessory apparatus near the search-coil. Hysteresis loops may be delineated with an accuracy of 2–3 per cent on remanence and about 1 per cent on coercivity, while the instrument is particularly valuable where rapid testing of a number of specimens is required.

Measurements on ferromagnetic materials. (*b*) *Mercury helix permeameter.* A permeameter operating on original lines has been described by Somaini[93], who claims that its fundamental principle ensures the constancy of its calibration with time, and that its readings are therefore accurately reproducible. The distinctive feature of his apparatus lies in the measurement of a magnetic flux by means of the pressure set up in a plane helix of mercury traversed by a mutually perpendicular current and flux. To form the mercury helix, two annular blocks of insulating material are placed side by side with a flat spiral of wire between them, the space so framed being filled with mercury. The helix is enclosed circumferentially by two concentric copper rings, which serve as electrodes for the passage of a radial current crossing mercury and wire alternately. The whole system is placed in the gap between two permalloy rings which form part of the return flux-path of the magnetizing solenoid, the helix being thus traversed by a flux perpendicular to its plane. In accordance with Ampère's law a pressure is set up along the spiral, proportional at any point to the product of the current density and the flux density, and by suitable design of the helix the total pressure difference between its inner and outer ends may be made proportional to the flux through the helix and hence to the magnetic p.d. across the gap. The pressure difference is measured by means of an alcohol manometer, the readings of which may thus be calibrated in terms of the gap p.d.

The magnetizing solenoid is fitted with massive permalloy yokes terminating in the rings which enclose the gap. Their reluctance is negligible in comparison

with that of the solenoid and the gap, and when there is no specimen in the permeameter we have therefore

$$0{\cdot}4\pi SI_0 = H_0 l + v_g,$$

where S, l are the number of turns and length of the solenoid; I_0, H_0 are the magnetizing current and field-strength in the solenoid, and v_g is the magnetic p.d. across the gap. From this we derive

$$H_0 = \frac{0{\cdot}4\pi S}{l} I_0 - \frac{v_g}{l} = \frac{0{\cdot}4\pi S}{l} I_0 - \frac{k.2\delta}{l},$$

where 2δ is the change of the manometer reading on reversing the magnetizing current, and k is a constant. On introducing into the solenoid a specimen, in the form of a rod or bar of uniform cross-section and of length equal to that of the solenoid, the flux across the helix tends to increase but is restored to its original value, as indicated by the manometer reading, by adjusting the magnetizing current to a lower value I_B. Then, equating the two values of the total flux,

$$\frac{0{\cdot}4\pi SI_0 - k.2\delta}{l} A = \frac{0{\cdot}4\pi SI_B - k.2\delta}{l} (A-a) + Ba,$$

where A, a are the cross-sections of the solenoid and specimen, and B is the flux density in the specimen. Hence

$$B = \frac{A}{a} \frac{0{\cdot}4\pi S}{l} (I_0 - I_B) + H,$$

where H is the field-strength corresponding to the magnetizing current I_B. Thus B and H are calculable from the two current readings and the constants of the solenoid and specimen, the manometer reading being involved only in the correcting term $k.2\delta/l$, which is practically negligible at flux densities above 13,000 G. but may amount to 20 per cent of the value of H in the region of high permeability.

Measurements on ferromagnetic materials. (*c*) *Magnetic hysteresis at low flux-densities.* The behaviour of magnetic materials at very low flux-densities has been the subject of a great deal of controversy. Lord Rayleigh's classical investigation in 1887 by a magnetometric method indicated that the hysteresis loop at flux-densities down to 30 G. is represented by a quadratic function of the field-strength, while the energy loss due to hysteresis is proportional to $B^3_{\max.}$. Subsequent a.c. measurements have disclosed an additional loss proportional to $B^2_{\max.}$, generally attributed to magnetic viscosity or after-effect, the exact nature of which is uncertain. Wittke[94], using a pendulum-switch, has investigated ballistically the time variation of flux-density after switching on a magnetizing current, and as a result of his analysis claims to find definite evidence for the existence of magnetic after-effect. Ellwood[95] has applied the resonating ballistic galvanometer, described in the 1935 *Report*, to the delineation of hysteresis loops on rings of compressed iron dust down to a $B_{\max.}$ of 2 G. The difficulties involved in this measurement may be realized from the fact that the 60 : 1 variation of $B_{\max.}$ covered by the measurements corresponded to more than 100,000 : 1 variation in the area of the loop, while the

remanence in the smallest loop was only $B_{max.}/1500$. To obtain the necessary precision, measurements were made on the deviations of the loop from the straight line joining its tips, by balancing the ballistic impulse for a complete reversal by means of an air-cored mutual inductance. The galvanometer deflection was built up by the method of repeated impulses described in the previous report, the necessary switching operations to select the favourable impulses to the galvanometer being performed automatically by means of a photocell actuated by a beam reflected from the galvanometer mirror. Ellwood's results show that Rayleigh's laws hold accurately down to the much lower flux densities used in his investigation and that after-effect does not influence the loss measured ballistically. Moreover, calculation of the third harmonic induced voltage by a Fourier analysis of the ballistic loops agreed with observed values using a.c. magnetization within experimental error, showing that any deviation of the dynamic from the static loop due to after-effect does not appreciably affect the third harmonic of the induced e.m.f.

Measurements on ferromagnetic materials. (*d*) *Incremental permeability measurements.* In recent years extensive use has been made of magnetic materials subjected to superposed d.c. and a.c. magnetizations—for example, in audio-frequency transformers and reactances. The best criterion of the behaviour of material in these conditions of magnetization is the incremental permeability, μ_Δ, which may be defined as the value of $\Delta B/\Delta H$, where ΔB is the total cyclic variation in flux-density resulting from the superposition of a total cyclic variation in magnetizing force ΔH on a steady magnetizing force, H. The measurement of μ_Δ is a matter of some difficulty owing to the complicating effects of the wave-form distortion caused by the non-linear relation between B and H. The available methods have been reviewed by Sims[96], who draws particular attention to the fact that when a polarizing magnetizing force is present the subsidiary hysteresis loop described by the material is unsymmetrical and even harmonics are present in the wave of either flux or magnetizing force. Methods based on the use of a rectifying commutator, such as are widely used with pure alternating magnetization, are thus unreliable when superposed magnetizations are present unless a commutator of special design is used giving adjustable time of contact as well as variable phase. The error due to the presence of even harmonics is likely to be negligible when the amplitude of the alternating magnetizing force is small, which is the most important practical case, but may become considerable for large amplitudes. Sims and Clay[97] have further carried out exhaustive tests on a sample of stalloy rings subjected to combined a.c. and d.c. magnetizations, which show conclusively the necessity for precise control of the conditions of measurement if reproducible results are to be obtained, and further indicate that ballistic measurements, besides taking no account of the effects of eddy currents, do not in general give reliable guidance as to the behaviour of a material in these circumstances.

Measurements on ferromagnetic materials. (*e*) *Changes in thermal energy accompanying changes in magnetization.* The increase in thermal energy which takes place when a substance is carried through a complete magnetic cycle is found to be the sum of a series of changes of thermal energy having both positive and negative

signs associated with the various portions of the cycle. Townsend[98] has investigated these changes in nickel using an adaptation of an apparatus previously employed by Ellwood. The specimen in the form of a wire was placed axially in a cylindrical calorimeter and magnetized by a solenoid. The temperature changes in the wire resulting from various changes in magnetizing force were measured by means of a battery of thermo-couples, and the corresponding energy changes were deduced by calibrating the thermo-couples by passing known currents along the specimen for known intervals of time.

It is of interest that the thermal changes in the end portions of the loop, when the magnetic changes are most nearly reversible, are of opposite signs in nickel and in carbon steel. No explanation of this effect has been given.

Measurements on feebly magnetic materials. (*a*) *Susceptibility measurements.* While no fundamental developments in the measurement of susceptibility are to be recorded, several minor modifications of well-established methods have recently been published. Thus both Kraeber[99] and Raychaudhuri[100] have discussed the testing of ferromagnetic powders. Kraeber describes a form of magnetic balance specially designed for measurements on powdered iron ores. By the use of specially shaped pole-shoes of vacuum-melted cobalt-steel an intense field is obtained which is almost uniform over a considerable distance in one direction and has an almost constant gradient over a considerable distance in a perpendicular direction. The readings are thus made largely independent of the shape of the specimen over a wide range. By the use of a very intense field (of the order of 5000 oersteds) the varying self-demagnetizing effect of different specimens is minimized and comparable results obtained. Raychaudhuri employed Bozorth's astatic magnetometer to test powdered ferrites having mass susceptibilities of the order of 0·1. In this instrument a specimen in the form of a rod, a few centimetres long, is mounted vertically in a magnetizing solenoid so that opposite poles act in the same sense on the astatic needle system. Owing to the small size of specimen required, it is very suitable for use with powder specimens. Raychaudhuri points out that the susceptibility of ferromagnetic powders is not a constant of the material but is greatly affected by the packing density, and has obtained curves connecting the susceptibility and the packing density.

A magnetic balance for susceptibility measurements on liquids or solids at temperatures up to 1000° C. has been described by Sosman and Austin[101]. It operates on Faraday's principle, the specimen being suspended in a field with a vertical gradient and the force on it being measured by attaching it to the arm of a sensitive balance. The special feature of their apparatus is the provision of two concentric solenoids, one of which provides a uniform field while the other applies a constant gradient. This gives great flexibility in operation and also makes it possible to examine the hysteresis loop of a material, the field-strength being varied cyclically while a constant gradient is maintained. The results obtainable on ferromagnetic powders, however, can only be qualitative owing to the self-demagnetizing effect of the specimen.

Measurements on feebly magnetic materials. (*b*) *Magnetostriction in bismuth*

crystals. The changes in dimensions produced by magnetization are extremely small even in ferromagnetic materials. In non-ferromagnetic substances they are so minute that in spite of the attempts of numerous investigators no positive results had been obtained until Kapitza[102] applied transient fields of the order of 3×10^5 oersteds to bismuth crystals and, by magnifying the resulting displacements hydraulically and optically, succeeded in measuring their magnetostrictive properties. Recently two independent investigations have been made into the magnetostrictive behaviour of bismuth crystals in steady fields of the order of 20,000 oersteds. Wolf and Goetz[103] mounted their specimens in a specially constructed framework placed inside the magnetizing solenoid from which it was screened by water-jackets. The change in length of the specimen relative to a neutral material rotated a small mirror reflecting a beam on to a thermoelectric relay of the type described in the 1934 *Report*, p. 324. A total magnification of 400,000 was thus obtained, allowing a strain of $2{\cdot}5 \times 10^{-8}$ in a specimen 10 cm. long to be read. Moreover a photographic record of the dimensional changes could be obtained. The apparatus was calibrated by raising the temperature of the specimen by 0·2° C. by means of a heating coil along its length. The temperature rise was measured by means of a platinum resistance coil wound on the specimen, and the expansion was then calculated from the known thermal properties of the material.

Shoenberg[104] obtained a slightly larger magnification by an adaptation of Kapitza's hydraulic method. The magnetostrictive displacement to be measured was applied through a system of fused quartz links (giving low thermal expansion) to a flexible diaphragm in the wall of an oil-chamber. Motion of the diaphragm drove the oil through a fine capillary, giving a magnification approximately equal to the ratio of the areas of diaphragm and capillary. The motion of the oil was observed by suspending in it very small particles of lamp-black, and watching the motion of individual particles through a microscope. Great care was taken to minimize thermal disturbances, but with the fullest precautions a steady drift still occurred. The effect of this was eliminated by plotting a curve of displacement with time for a particular particle before and after the magnetostrictive displacement was applied. The apparatus was calibrated by Kapitza's method by applying a known displacement to the diaphragm by bending a steel rod, the elastic properties of which had been independently determined. An accuracy of about 10 per cent is claimed on the magnetostrictive moduli, and the results obtained agreed well with those of Kapitza for much stronger fields. The method can be applied, however, as Kapitza's cannot conveniently, to measurements of transverse magnetostriction.

References

(84) Yensen, T. D. and Ziegler, N. A. *Trans. Amer. Soc. Metals*, **23**, 556 (1935).
(85) McKeehan, L. W. *J. Franklin Inst.* **197**, 583, 787 (1924).
(86) Von Auwers, O. and Neumann, H. *Wiss. Veröff. Siemens-Werk.* **14** (2), 93 (1935).
(87) Kelsall, G. A. *Physics*, **5**, 169 (1934).
(88) Dillinger, J. F. and Bozorth, R. M. *Physics*, **6**, 279 (1935).
(89) Goss, N. F. *Trans. Amer. Soc. Metals*, **23**, 511 (1935).
(90) Sixtus, K. J. *Physics*, **6**, 105 (1935).

(91) Bozorth, R. M. *Trans. Amer. Soc. Metals*, **23**, 1107 (1935).
(92) Stäblein, F. and Steinitz, R. *Arch. Eisenhüttenw.* **8**, 549 (1935).
(93) Somaini, G. *Elettrotecnica*, **22**, 588 (1935).
(94) Wittke, H. *Ann. Phys.*, Lpz., **23**, 442 (1935).
(95) Ellwood, W. B. *Physics*, **6**, 215 (1935).
(96) Sims, L. G. A. *Wireless Engr*, **12**, 8, 65 (1935).
(97) Sims, L. G. A. and Clay, D. L. *Wireless Engr*, **12**, 238, 312.
(98) Townsend, A. *Phys. Rev.* **47**, 306 (1935).
(99) Kraeber, L. *Naturwissenschaften*, **23**, 160 (1935).
(100) Raychaudhuri, D. P. *Indian J. Phys.* **9**, 415 (1935).
(101) Sosman, R. B. and Austin, J. B. *J. Wash. Acad. Sci.* **25**, 15 (1935).
(102) Kapitza, P. *Proc. roy. Soc.* A, **135**, 537 (1932).
(103) Wolf, A. and Goetz, A. *Phys. Rev.* **46**, 1095 (1934).
(104) Shoenberg, D. *Proc. roy. Soc.* A, **150**, 619 (1935).

§ 4. PIEZO-ELECTRICITY

By P. Vigoureux, M.Sc., A.C.G.I., D.I.C., A.M.I.E.E.

The use of piezo-electric material, in particular of quartz, for the measurement, standardization, and control of radio frequencies, has been dealt with in the *Report on Progress in Physics* for 1934 (p. 336), and the work of Cady[105], Dye[106], Giebe and Scheibe[107], Pierce[108] and others has been reviewed. Most of the work done recently on quartz vibrators has been directed to the production of vibrators of negligible temperature coefficient of frequency.

In ordinary resonators the temperature coefficient of the natural frequency is by no means negligible, and can reach nearly 1 part in 10,000 per °C. When high accuracy is aimed at, it is therefore essential to arrange for temperature regulation. Even with the best temperature regulation available, however, it would be impossible to reach a constancy of frequency of 1 in 10^8 if the temperature coefficient exceeded 2 or 3 parts in a million per °C. The temperature coefficient of frequency of quartz depends on the orientation which the resonator bears to the axes of the crystal from which it was cut, and it has been found possible to reduce this coefficient to zero by a suitable choice of orientation. Much of the recent work on quartz has been directed to that end.

Resonators in the form of rods of rectangular section have been studied by Giebe and Scheibe[107]. The rods always had their lengths perpendicular to the optic or principal axis Z of the crystal; in some cases the length was along the electric axis X, in other cases along the Y axis. In each case excitation is effected by means of small electrodes placed symmetrically a small distance from the faces perpendicular to the electric axis. The temperature coefficient is in general small, lying between 0 and -3×10^{-6} per °C., provided the breadth and the thickness are small compared with the length, say a quarter of the length or less, but between these values there is considerable and irregular variation. Measurements with shorter rods and measurements on the overtones of long rods, gave numerically higher coefficients, in one case nearly -20×10^{-6} per °C. All these irregularities are explained by the coupling between the longitudinal vibrations in the direction

of the length with the longitudinal lateral vibrations and the torsional vibrations. whose frequencies are dependent also on lateral dimensions. The effect of coupling on the frequency constant of cylinders and tubes, and of rods of rectangular section, has been studied in detail by Giebe and Scheibe[109] and by Giebe and Blechschmidt[110]. Whenever the ratio of these frequencies is nearly integral, the vibration is not pure, and the frequency constant as well as the temperature coefficient show appreciable departures from the predicted values. In this respect the rods whose length is along the electric axis are superior to the others, because in this case the applied electric potential does not tend to excite torsional vibrations. For this type of cut the temperature coefficient is on the whole smaller than for the rods whose length is along the Y axis. Measurements with rods of square section, $(y=z)$, showed that, to a very rough approximation, the temperature coefficient is a linear function of ky/x, where k is the order of the overtone, y the lateral dimension, and x the length, of the resonator. For small values of the parameter, i.e. for long rods, the coefficient is negative, and of the order of 10^{-6}; it is nearly zero when the parameter is equal to 0·27, and it is positive for larger values. This law of variation makes it possible to choose the dimensions of the resonators so as to obtain negligible temperature coefficients. The coefficient varies slightly with temperature, and in general it vanishes at a particular temperature only. Resonators having a vanishingly small coefficient at about 35° C. have been designed on this principle, and are in use in the quartz clocks of the Physikalisch-Technische Reichsanstalt[111].

Longitudinal vibrations can be excited piezo-electrically in rods whose lengths are in directions other than X or Y. The characteristics of rods cut in the YZ plane have been investigated by many workers, notably by Matsumura and Kanzaki[112] and by Bechmann[113]. Although the tensile strain produced by an electric force directed along X is at right angles to Z, still the component in the direction of the rod is sufficient to excite vibrations in all rods cut in the YZ plane, except possibly those which are oriented along Z itself. In order to obtain the temperature coefficient of frequency of the actual longitudinal vibrations, it is essential to keep the lateral dimensions of the rods small compared with the length, otherwise the couplings studied by Giebe and Scheibe can vitiate the results. Bechmann[113] and Matsumura and Kanzaki[112] made measurements on a large number of rods inclined at various angles with the optic axis Z, and with their lengths in the YZ plane, and they found that the temperature coefficient is vanishingly small for an angle of about 95°. As this orientation differs by only 5° from the orientation for which the piezo-electric effect is a maximum, the rod can be easily excited, and its general performance is very good.*

The natural frequency of longitudinal vibration of a rod of length l is equal to $1/2l\sqrt{(s\rho)}$, where ρ is the density of quartz and s the elastic constant (the reciprocal

* The values obtained for the angles depend on the convention used for the axes. Here we follow Voigt's convention: the positive Y comes out through one of the three faces R—faces of the direct rhombohedron—which form a pyramid pointing in the direction of positive Z, and the direction of positive X is chosen such as to yield a right-handed system of co-ordinates.
The angle referred to above is measured from positive Z towards positive Y.

of Young's modulus) in the direction of the length of the rod. The temperature coefficient of frequency therefore depends on the temperature coefficient of expansion and on the temperature coefficient of the elastic constant in the particular direction. These quantities can be related to the two coefficients of expansion, viz. α_z along Z and α_x perpendicular to Z, which are known with considerable accuracy, and to the six coefficients of the elastic constants, viz. $\delta_{s_{11}}$, $\delta_{s_{33}}$, $\delta_{s_{44}}$, $\delta_{s_{12}}$, $\delta_{s_{13}}$, and $\delta_{s_{14}}$ by formulae which contain also the direction cosines λ, μ, ν, or two angles χ and ψ. By measuring the natural frequencies of a number of rods of different orientations at various temperatures, data are obtained from which the elastic constants s_{11}, s_{33}, s_{44}, s_{12}, s_{13}, s_{14}, and their thermal coefficients, can be evaluated. This evaluation has been performed by Bechmann[113] from data obtained on a large number of rods. Once the values of the individual coefficients are known, it is easy to calculate the temperature coefficients of Young's modulus and of frequency of a rod of any orientation.

The smallness of the temperature coefficient of frequency of modes of vibration which have no coupling to other modes is also illustrated in the ring oscillator constructed by Dye[114]. This oscillator consists of a cylindrical ring of quartz with its axis parallel to the optic axis. Six pairs of electrodes disposed along the circumference are used to excite the ring, alternate pairs being cross-connected, so that the applied potential causes expansion or contraction of all six segments of the ring at the same time. A mode of vibration is thus obtained for which the ring expands and contracts radially as a whole, the frequency being approximately equal to $1/2\pi a\sqrt{(s_{11}\rho)}$, in which a is the mean radius of the ring, ρ its density, and s_{11} the elastic constant or reciprocal of Young's modulus in a direction perpendicular to the optic axis. The actual frequency is 20,000·00 c./sec. The dimensions of the ring are such that there is no other mode of vibration present. Thus there is no compensation due to the influence of other modes of vibration of different temperature coefficient, and the temperature coefficient of the oscillator approximates to that of pure longitudinal vibrations in quartz. The value found by measurement is $-2{\cdot}5$ parts in a million per 1° C. rise in temperature, between the temperatures 15 and 20° C.

When the rods of Giebe and Scheibe are employed for frequencies exceeding 1 megacycle per sec., it becomes impossible to keep the ratio of the lateral dimension to the length, multiplied by the order of the overtone utilized, small enough to avoid coupling and to secure a small temperature coefficient of frequency. For frequencies exceeding that value it is therefore usual to employ rectangular plates or disks. If the normal to the plate is in the direction of the electric axis, the vibration is longitudinal, whereas if the normal is parallel to the Y axis a shear vibration is produced. In either case the electrodes are metal plates applied to the faces of the quartz plate, or a short distance away from the faces, and the electric field is in the direction of the normal to the plate. It has been shown by Koga[115] that the natural frequency of a thin plate of thickness t is $\frac{1}{2}t^{-1}\sqrt{(c'_{ik}/\rho)}$, where c'_{ik} is the elastic modulus corresponding to the vibration. For a plate normal to X this modulus is c_{11}, and for a plate normal to Y it is c_{66}. The figure which, when divided

by the thickness in millimetres, gives the frequency in kilocycles per sec., is called the *frequency constant*.

Just as for the rods of Giebe and Scheibe, the frequency constant of a plate is apt to suffer large alterations when there is coupling between the main vibration, in this case the thickness vibration, and the overtones of modes whose frequency is governed wholly or in part by the other dimensions of the plate. The effects of coupling have been studied by Lack[116], and more recently by Mason[117], whose theory succeeds in explaining the complicated changes which take place in the natural frequencies of rectangular plates perpendicular to the electric axis, when the ratio of the lateral dimensions is changed, the thickness remaining unaltered. He has, moreover, shown that the coupling, and therefore also the strength of subsidiary resonances, can be greatly reduced by the use of plates in which the longer dimension makes an angle of $-18{\cdot}5°$ with Y.

Plates cut perpendicular to X or to Y labour under the disadvantage of relatively large temperature coefficients, viz. about -20 and $+80$ parts in a million respectively per 1° C. These large temperature coefficients render the plates unsuitable for highly accurate frequency standardization. It has been found, however, that it is possible to obtain very small thermal coefficients of frequency by cutting plates in such a way that their normals are inclined to two axes, X and Z, or Y and Z, and perpendicular to the remaining axis. These plates have been studied by Koga[118], by Lack, Willard and Fair[119], and by Bechmann[113]. The plates which contain the Y axis, i.e. the plates whose normals have direction cosines $\sin\theta$, 0, $\cos\theta$, do not readily lend themselves to the production of resonators of zero temperature coefficient. When θ is 90°, the temperature coefficient of frequency is, as mentioned above, about -20 parts in a million; and there is a single thickness resonance, the mode of vibration being longitudinal. For other values of θ between 0 and 180° there are, besides the longitudinal mode, two shear modes of vibration[113], the frequency constants lying between 1600 and 2500 kc.-mm. Experiment shows that the temperature coefficient of the longitudinal mode is negative for all values of θ, as is also that of the first shear mode; but the temperature coefficient of the second shear mode changes sign somewhere between 50 and 70°. As, however, it is disadvantageous to have three main resonances not very far apart, and as the two shear modes are apt to be accompanied by subsidiary resonances[(9)], the $\sin\theta$, 0, $\cos\theta$ plates are not as a rule employed where a low temperature coefficient is required.

Much better results can be obtained from plates containing the X axis, i.e. plates whose direction cosines are 0, $\sin\theta$, $\cos\theta$. Piezo-electric excitation of these plates gives rise to a single shear mode of vibration, whose frequency constant varies between 1650 and 2500 kc.-mm., depending on the value of θ. The temperature coefficient of frequency varies with θ between $+80$ and -80 parts in a million, and it vanishes for values of θ equal to 42 and 125°. Of these two orientations the latter is the more advantageous, because, as shown by Lack[119], the piezo-electric constant d'_{26}, which is responsible for the excitation of the plate, is higher for the 125° orientation than for the other orientation, and the elastic

modulus c'_{56} is very small for the 125° orientation. As a result of the smallness of this modulus, the coupling between vibrations governed by the thickness and vibrations governed by the other dimensions of the plate is negligible, and a pure mode is obtained. Thus the abrupt changes of frequency which usually occur when the temperature of plates perpendicular to Y is altered gradually are eliminated by using this orientation. At the same time, as the thermal coefficient of frequency of the plate is negligible over a large range of temperature, and as plates of this type, vibrating in shear, can be rigidly clamped at the periphery or at the corners, depending on the shape, they constitute very robust and permanent resonators.

The possibilities afforded by the use of high-precision piezo-electric vibrators are illustrated by data recently obtained with the three quartz oscillator clocks maintained in operation at the Physikalisch-Technische Reichsanstalt[120]. The clocks are compared with one another and with observatory time-signals in such a manner that at ten-day intervals their rates relative to astronomical time are determined with a precision of ± 0·001 sec. per day. During the month of June 1935 all three clocks showed an increase in rate of 0·004 sec. per day, and preserved this new rate until October 1935 when they showed a more gradual decrease in rate. Since the clocks are independent units and are not affected in equal degrees by variations in physical conditions such as temperature, it seems unlikely that they should all show such equal variations at the same time; and the evidence indicates that the changes are results of variations of astronomical time, i.e. of the period of revolution of the earth. It was already suspected from astronomical observations of other cosmic movements extended over the last two centuries, that small variations of the period of the earth's rotation existed.

References

(105) Cady, W. G. *Proc. Inst. Radio Engrs*, N.Y., **10**, 83 (1922).
(106) Dye, W. D. *Proc. phys. Soc.* **38**, 399 (1926).
(107) Giebe, E. and Scheibe, A. *Hochfrequenztech. u. Elektroakust.* **41**, 83 (1933).
(108) Pierce, G. W. *Proc. Amer. Acad. Arts Sci.* **59**, 81 (1923).
(109) Giebe, E. and Scheibe, A. *Ann. Phys.*, Lpz., **9**, 93 (1931).
(110) Giebe, E. and Blechschmidt, E. *Ann. Phys.*, Lpz., **18**, 417 (1933).
(111) Scheibe, A. and Adelsberger, U. *Hochfrequenztech. u. Elektroakust.* **43**, 37 (1934).
(112) Matsumura, S. and Kanzaki, S. *Rep. Radio Res. Japan*, **2**, 35 (1932).
(113) Bechmann, R. *Hochfrequenztech. u. Elektroakust.* **44**, 145 (1934).
(114) Essen, L. *Proc. roy. Soc.* A, **155**, 498 (1936).
(115) Koga, I. *Rep. Radio Res. Japan*, **2**, 157 (1932).
(116) Lack, F. R. *Bell Syst. Tech. J.* **8**, 515 (1929).
(117) Mason, W. P. *Bell Syst. Tech. J.* **13**, 405 (1934).
(118) Koga, I. *Rep. Radio Res. Japan*, **4**, 61 (1934).
(119) Lack, Willard and Fair. *Bell Syst. Tech. J.* **13**, 453 (1934).
(120) Scheibe, A. and Adelsberger, U. *Phys. Z.* **37**, 185 (1936).

ELECTRICAL METHODS OF COUNTING

By C. E. WYNN-WILLIAMS, Ph.D.

§ 1. INTRODUCTION

During the last few years much attention has been given to problems in nuclear physics connected with artificial disintegration and induced radioactivity. As in all problems connected with radioactive transformations, owing to the random nature of the emission of the electrified particles, the counting of which forms the basis of the experiments, the observations can only be relied upon if the individual counts comprise numbers sufficiently large to make the statistical variations small. It has therefore become increasingly important to be able to make frequent counts of particles at high counting speeds. For this reason, the scintillation screen, which was much employed in earlier work, is now seldom used. In place of it, new high speed electrical counting methods have been developed, and the recording of particles by these methods is no longer an experimental curiosity, but an established and important laboratory technique.

In principle, the methods consist in causing the particles to be detected to produce minute current pulses within a small ionization chamber, or alternatively, rather larger current pulses in a Geiger-Muller counter. These pulses are then passed through a valve amplifier—the amount of amplification required being greater in the former instance—until they are sufficiently magnified to operate some sort of recording device. This device may be a telephone receiver for counting by ear, an oscillograph for obtaining a permanent photographic record of the entry of the particles into the chamber, and also possibly a measure of the ionization produced by each particle, or finally it may be an automatic thyratron counter capable of counting with accuracy several thousand random particles per minute. In the last instance, auxiliary electromechanical devices may be incorporated which will carry out, entirely automatically, all the operations necessary for taking a series of timed counts, even, if desired, to printing the experimental results in plain figures on a paper tape. A good deal of tedious routine observational work may thus be avoided, and statistical results collected very rapidly.

In the various papers bearing on the subject by different writers, numerous circuit diagrams have been published. To the physicist who is not specializing in this particular work, and who has not the time to go into the technicalities, the multiplicity of apparently different arrangements of valves and thyratrons is naturally confusing. It is the object of the writer therefore, to give a rather simple and general account of the modern electric counting methods, and the experimental difficulties peculiar to them. Diagrams are given of some typical counting circuits; lack of space, however, prevents a detailed discussion of the experimental arrangement. This article should not therefore be regarded as providing in itself all the

information necessary for setting up the circuits. This information, if desired, must be obtained from the original published papers, a list of some of which is appended.

§ 2. DETECTION OF PARTICLES

The particles to be detected must first be made to produce potential pulses of sufficient magnitude to operate a valve amplifier. For α-particles or protons, a small ionization chamber of low capacity and high insulation may be used. The positive ions from a few millimetres of the particle track in air at ordinary pressure are driven quickly by an electric field of about 1000 V./cm. on to the collecting electrode, causing the potential of the latter to rise sharply by about 10^{-4} V. If slow neutrons are to be detected, the chamber may be lined, or the electrodes covered, with boron or lithium. Alternatively, Goldhaber, and Fermi, have used chambers containing boron trifluoride gas instead of air. The ionization is then produced by the α-particles emitted from the boron or lithium under neutron bombardment. Fast neutrons may be detected either by slowing them down by passage through, say, paraffin, contained without or within a boron or lithium chamber, or directly by means of the helium pressure chamber method, described by Oliphant. In this latter method, the artificial α-particles produced by collisions of fast neutrons with helium atoms cause ionization of the helium contained in the chamber.

In some radioactive transformations the particles emitted are fast positive or negative electrons, or β-particles, of much lower ionizing power than α-particles or protons. To detect these, it is usual to employ a Geiger-Muller tube counter. Compared with the ionization chamber such a counter suffers from the drawback of requiring a steady high potential of about one or two thousand volts, and of possessing a much higher natural effect. Because of the multiplication of the primary ionization current by collision, however, a very much simpler amplifier, usually requiring only about half the number of stages necessary with an ionization chamber, may be employed.

§ 3. AMPLIFICATION OF PULSES

The practicability of detecting α-particles by linear amplification of the pure ionization current was first demonstrated by Greinacher[1], who used for the purpose a transformer-coupled amplifier. Since then the method has been further developed and used by other workers. In this section, an explanation is given of the process of amplification in a "resistance-capacity" type of amplifier, such as is now generally used for counting work. Figure 1 shows only the essential details of an amplifier in the simplest possible form. Geiger-Muller circuit arrangements are discussed later, in § 9.

First valve stage. When a particle enters the chamber, a trail of ions is formed in the gas. The positive ions are driven by the electric field on to the collecting electrode, which is connected directly with the grid of the first valve. A positive rise in the potential of the grid* then takes place as the ions are collected, followed

* In the explanation of this and of other circuits given in the article, all electrode potentials are stated with reference to that of the valve cathodes, which are assumed to be at zero or "earth" potential.

by a fall as the accumulated charge leaks away and the grid returns to its normal floating potential. The shape and magnitude of this voltage pulse obviously depends upon the number, and the rate of collection, of the ions, and the magnitude of the natural capacity and leakage of the grid-chamber system. At this stage it is sufficient merely to point out that in order to secure a large voltage pulse, the rate of collection should be high, and the capacity and leakage low.

The resulting positive rise and fall of grid potential causes the anode current of the first valve to vary in a similar manner. Since this current passes through the anode resistance R_1, the current variation will produce a similar variation (i.e. a positive rise and fall) in the voltage developed across this resistance. As the upper end of the resistance is maintained at a fixed positive potential V, the actual voltage variation must occur at the first valve anode. The potential of the latter therefore

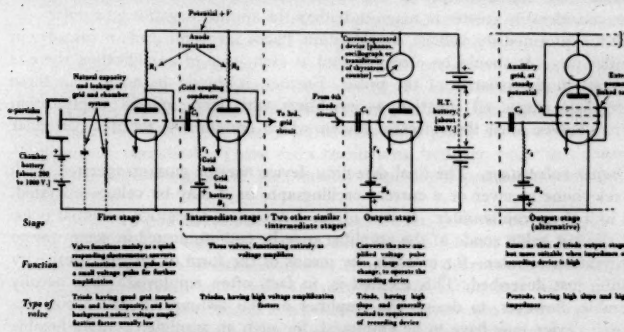

Potential +V
Anode resistance
Natural capacity and leakage of grid and chamber system
α-particles
Chamber battery (about 200 to 1000 V.)
Grid coupling condenser
To ht. grid circuit
Grid leak
Grid bias battery
From ht. anode circuit
H.T. battery (about 200 V.)
To H.T. battery
Auxiliary grid, at steady potential
Extra grid permanently linked to cathode
Stage
First stage
Intermediate stage
Two other similar intermediate stages
Output stage
Output stage (alternative)
Function
Valve functions mainly as a quick responding discriminator; converts the ionization current pulse into a small voltage pulse for further amplification
Three valves, functioning entirely as voltage amplifiers
Converts the now amplified voltage pulse into a large current change, to operate the recording device
Same as that of triode output stage, but more suitable when impedance of recording device is high
Type of valve
Triode having good grid amplification and low capacity, and low background noise; voltage amplification factor usually low
Triodes, having high voltage amplification factors
Triode, having high slope and generally suited to requirements of recording device
Pentode, having high slope and high impedance

Figure 1. Simplified diagram of ionization chamber and amplifier system.

falls as the grid potential *rises*, and vice versa. Thus the sudden production of ions in the chamber results in the application of a small positive pulse to the grid, and this, in turn, causes a *negative* pulse of similar shape but greater magnitude to be superimposed upon the potential of the anode. It will be clear, therefore, that the first valve functions primarily as an electrometer of quick response, and converts the small ionization *current* pulses into *voltage* pulses suitable for further amplification. Since this is one of the most important points to consider in choosing the type of valve for the first stage, it may be necessary to sacrifice a high voltage amplification factor in favour of other more desirable characteristics.

Intermediate valve stages. Further amplification is necessary before recording devices can be operated. The negative anode potential pulse from the first valve must be communicated in some way to the grid of the next valve. If a direct connexion were made, the two electrodes concerned would always be at the same relative potential. For correct operation as an amplifier, however, the grid of a valve must be maintained at some suitable mean negative potential (chosen by reference

to the valve characteristic curve), and upon this potential must be superimposed any positive or negative voltage variation it is desired to amplify. For this reason the connexion between the two stages is taken through the grid coupling condenser C_1, and the grid of the second valve is maintained at the necessary mean negative potential by the grid-bias battery B_1, through the high resistance grid leak r_1.

The negative anode potential pulse from the first valve is now communicated, through condenser C_1, to the grid of the second valve. The voltage amplification process which follows is similar to that described for the first valve, with the difference that all polarities of the pulses are reversed. The *negative* pulse in the potential of the second valve grid causes a momentary *reduction* in the anode current. Since this anode current traverses R_2, the upper end of which is at a fixed potential $+V$, the anode of the second valve is subjected to a *positive* potential pulse, considerably greater in magnitude than the applied negative grid pulse.

Other intermediate voltage amplification stages are connected in cascade, in a similar way. It should be observed that at each stage of amplification there is a reversal in the polarity of the pulse. Further, it should be noted that these intermediate stages all function as amplifiers with large voltage amplification factors. Valves must therefore be chosen which are suitable for this particular purpose.

Output valve stage. The final detecting device may be current-operated, such as a telephone receiver or a current oscillograph, or it may be voltage-operated, such as a thyratron counter. If it is of the latter type, a positive potential pulse from the last valve anode of the amplifier may be communicated in some way to the thyratron counter—for example, by means of the form of resistance-capacity coupling just described. This method is, in fact, often employed. It is usually preferable, however, to design the amplifier on the assumption that a current-operated device may have to be employed, for such an amplifier is more flexible in use. Voltage-operated devices may then always be operated from it by means of a transformer through which the output valve anode current passes (see figure 6*a*, § 13) or from the voltage set up across a resistance or other impedance through which the anode current passes (figure 6*b*).

The most suitable output valve is therefore one in which a large change of anode current is produced by a small change of grid potential—that is, a valve having a large "slope", or "mutual conductance". It should be remembered, however, that the inclusion of a resistance or impedance in the anode circuit of a valve modifies the form of the {anode current, grid voltage} characteristic curve. If therefore, the current-operated device possesses high inductance, it may be necessary, in order to avoid distortion of the shape of the current pulse which passes through it, to use an output valve having an impedance high compared with that of the device.

Advantage of pentode output valve. For a triode of a given type, high impedance is only obtained at the expense of slope. As both of these qualities are needed, it is usually preferable to use a pentode in the output stage, especially when the

recording device is a current-operated oscillograph. Unlike triodes, pentodes combine the advantages of both high impedance and high slope, or mutual conductance. Although a pentode possesses three grids, one is already connected with the cathode inside the valve envelope itself. The other extra grid merely requires maintaining at a steady potential rather less than that of the anode, exactly as does that of a screen-grid valve. Thus, the inclusion of a pentode in the amplifier is not a great complication (see figure 1).

Choice of circuit arrangements. In the type of amplifier just described, resistance-capacity coupling is used. There are, however, two other very well-known methods of coupling, transformer coupling and battery coupling, and both of these have been used for counting purposes. Ortner and Stetter[2] have shown that, of the three methods, only the one employing battery coupling is free from distortion. They therefore used this method in their work, in spite of certain disadvantages. If, however, the time constant—that is, the product of r and C in figure 1—is made as large as practical considerations will permit,* the distortion of a pulse is not sufficiently great to be troublesome. Owing to their greater convenience, therefore, practically all counting amplifiers are now of the resistance-capacity type. In practice, C would be about 0·1 μF., and r about $\frac{1}{4}$ to MΩ. Under these conditions, a pulse becomes distorted into a pulse followed by one or two oscillations of much smaller amplitude. These extra oscillations, however, occur at a frequency so low that an oscillograph of high natural frequency is practically unaffected by them.

Use of screen-grid valves instead of triodes. Screen-grid valves possess the great advantage of having very high voltage amplification factors. They may be used instead of triodes in a resistance-capacity amplifier, if provision is made to supply the extra grid of each valve with a steady potential rather less than that of the anode. For efficient operation, it is usually necessary to use much higher values of anode resistances, and anode voltages, than for triodes. Johnson and Neitzert[3] have, however, described an amplifier consisting of four screen-grid valves and one pentode, operating at low anode potentials, for which a very low background noise level is claimed. Abroad, particularly in America, counting amplifiers are more usually constructed with screen-grid valves. In this country, however, owing to the fact that triodes are available having amplification factors of the order of 75–100, the tendency has been to simplify matters as far as possible by using triodes instead of screen-grid valves. Under normal circumstances, four triodes and one pentode provide adequate magnification for counting α-particles or protons. In spite of the higher amplification factors of screen-grid valves, the margin of amplification available is usually insufficient to enable fewer stages of screen-grid amplification to be employed than of triode amplification.

* If, under abnormal circumstances, a positive pulse greater in magnitude than the negative grid bias is applied to the grid, "grid current" flows, and the grid may subsequently be left at a negative potential sufficiently high to prevent the valve from functioning. If the time constant of the circuit is too high, this state will persist for an inconveniently long time. This phenomenon (which often occurs with the *first* valve owing to the highly insulated grid), is known as "grid-choking".

§ 4. FILTER CIRCUITS

In order to avoid instability, the simple circuit of figure 1 must be modified. During the amplification of a pulse, each anode current is subjected to a fluctuation, either positive or negative, that of the last valve being by far the greatest. Suppose now the internal resistance of the high-tension battery or other source of current is not negligible. The voltage V between the battery terminals will fluctuate, and potential pulses will be fed back through the anode resistances to the anodes of all the other valves. That arriving at the first valve anode will be communicated through C_1 to the grid of the next valve, and will be amplified by this, and by the remaining stages. In turn, this amplified pulse will produce another battery voltage fluctuation. Thus, it will be seen that an amplifier may be unstable on account of

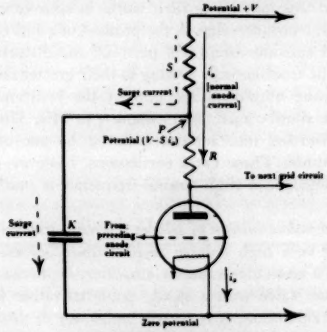


Figure 2. Filter circuit.

this "back-coupling", the degree of instability depending, amongst other things, upon the internal resistance or impedance of the source of anode current. With such an amplifier, low-frequency oscillations or periodic fluctuations of the anode currents will be experienced, the periodicity being dependent upon the values of the capacities, resistances, voltages and types of valves used in the circuit.

Instability of this kind may be prevented either by providing each stage with an independent anode current supply, or by trapping the "feed-back" pulses, so that they do not reach the anodes of the valves. In practice, complete prevention by either method alone is expensive and inconvenient, so a compromise is usually made between the two. In the first place, two independent sources of anode current are used, of reasonably low internal impedance, and the supplies to the various stages are divided amongst these two sources. Secondly, a filter or de-coupling circuit is inserted in each anode circuit. An example of such a circuit is shown in figure 2. This consists merely of a fairly high resistance S and a large condenser K. Normally, the point P (previously connected directly to the battery) will be at some

steady potential ($V - Si_a$), where i_a is the anode current. If an unwanted potential pulse is superimposed upon the battery voltage V, the pulse, instead of passing through R to the anode, is partly diverted through the low impedance path presented by the condenser K. Thus the magnitude of the potential fluctuation at the anode is considerably reduced. Alternatively, the charged condenser K may be regarded as a reservoir, or an independent subsidiary battery for this particular stage. As current is drawn from it, the condenser is being continually recharged from the main supply through S. It will be clear that the filter circuit serves also the purpose of smoothing out irregularities which may occur in the high-tension battery, or other source of supply.

In addition to preventing back-coupling, the filter is valuable in that it tends to keep out of the amplifying system the high-frequency pulses or oscillations which are often picked up by induction from other electrical apparatus or machinery. It may even be necessary to include filters in other parts of the amplifier circuit to protect them from the high potential surges which are sometimes carried along the laboratory power leads when high voltage apparatus is being used in the neighbourhood.

Obviously, the larger K and S are made, the more efficient is the filter. The practical limit of about 4–8 μF. for K is set by consideration of expense and bulk The limit for S is, however, set by consideration of the current to be passed and the maximum permissible voltage drop across S. If the current is large, S may have to be a low-resistance iron core choke. On the other hand, when it is merely desired to trap very high-frequency pulses or surges, S may be simply an air core choke of a few hundred turns, while K may be a fraction of a microfarad, owing to the very low impedance presented by quite a small condenser to high-frequency currents.

Several examples of filters are shown in figure 3. The ordinary type just described is seen in each of the anode circuits, and also in the circuit of the pentode auxiliary grid. A double stage filter is seen in the lead to the high potential chamber electrode. Inductance-capacity filters are included in the output circuits of the two cuprous oxide rectifiers, one a single, and the other a double, filter. These smooth out the ripple at 100 c./sec. from the rectified current. The 30,000-ohm resistances connected across the rectifiers serve merely to stabilize the load, and to prevent undue rise of voltage under abnormal conditions.

§ 5. ELECTRICAL SCREENING

The provision of electrical screening is far more important in the case of an amplifier than an electrometer. With the latter instrument it is usually sufficient to screen each part of the circuit, and to join the various sections by means of wires to a common earth point. An amplifying system containing a valve with a free grid, however, is very sensitive to electromagnetic surges and oscillations which may be emitted from neighbouring electrical apparatus or machinery, which would not affect an electrometer. It is therefore essential to provide the amplifier with

screening which will keep out electromagnetic waves, as distinct from electrostatic fields. Ideally, the whole system—amplifier, chamber, batteries and recording instrument—should be enclosed in one electrically continuous metal case. This is usually impracticable, so the chamber and first valve, the remainder of the amplifier, and the batteries or rectifiers, are each enclosed in a metal box, and the separate boxes are linked together by well-screened flexible leads. The recording instrument, as a rule, is situated at some distance from the amplifier; this instrument need not be screened provided the lead to it is well screened.

In designing the screening, the main point to bear in mind is that there should be electrical continuity in a direction parallel with the enclosed conductors. A narrow gap between the adjacent ends of two screening tubes, for example, should be bridged over by a *very short low-resistance* wire link, parallel with the conductors screened by the tubes. Screening covering on flexible wires should consist of close woven low-resistance wire, with the strands approximately parallel to the run of the wire. A covering of metal tape or wire wound in a spiral round the conductors, even with adjacent turns in contact, affords practically no protection. It should be remembered that unscreened insulated wires entering the screening boxes may behave as small aerials, and carry in high-frequency oscillations or surges. Filter circuits may therefore have to be inserted in leads entering or leaving the boxes, unless such leads are already well screened. An example of this is seen in the two 0·1 μF. condensers connected between the two alternating-current supply wires and the screen in figure 3.

With regard to the actual connexion to earth, a multiplicity of wires must be avoided. Only one point of the system should be connected, by a single fairly short wire, to a satisfactory earthing point in the laboratory, not used as a common earth for other electrical apparatus.

Inter-stage screening. Usually, it is also necessary, for stability, to provide screening partitions within the amplifier. If this is not done, the amplified pulses of anode current or pulses of anode potential, may, by electromagnetic or electrostatic induction, produce potential pulses in the grid circuit of this or of other valves in the amplifier. Thus, potential pulses may be fed back, and continuous oscillation may occur. In order to avoid this, the individual valves, coupling circuits, and filter circuits are screened from one another, and all leads within the amplifier, particularly those of the grid and anode circuits, are kept as short and direct as possible.

§ 6. PRACTICAL DETAILS OF AMPLIFIER

In figure 3 a circuit diagram is given of an amplifier designed and used at the Cavendish Laboratory. Most of the diagram is self-explanatory after what has already been said, so attention is merely directed to a few points of special note.

Current supplies. As far as possible, current is derived from the alternating-current mains. The high-tension current, at about 250 V., is supplied through smoothing circuits, from either of two metal-oxide rectifiers. These are provided with separate transformers, to avoid back-coupling. Chamber potential may

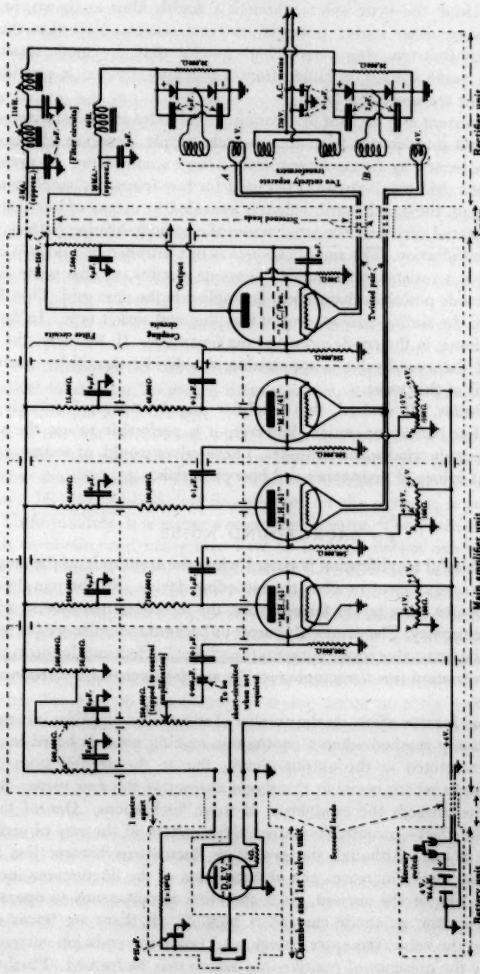


Figure 3. Detailed diagram of ionization chamber and amplifier system. All "earths" in this figure represent direct internal connexions to the nearest point of the metal screening boxes, represented by the thin broken lines. There is only one direct external connexion (not shown) between the the screen system and the laboratory earth wire.

also be derived from the same source, through a double filter as shown, or alternatively from some other higher potential external source. Note that the filter resistances in the first two stages are of higher value than the anode resistances, and that a low anode voltage, obtained from a potential divider, is provided for the first valve, for reasons given later.

Alternating current may be used for heating the cathodes of the last three valves. In order to avoid interference from the 50-cycle supply, however, the first two valves should be heated by direct current, from a large accumulator. In preference to the "automatic grid bias" circuit often used for low-frequency amplifiers, grid bias for the second, third and fourth valves is provided by means of low-resistance wire-wound potential dividers connected across the cathode-heating battery.

Control of amplification. The amplification may be controlled by using a potential divider instead of a resistance in one of the anode circuits, so that some desired fraction of the anode potential change may be applied to the next grid. This divider should not be of the sliding contact, but of the plug and socket type. In figure 3, this divider is shown in the anode circuit of the first valve. If, however, the background noise of the *second* valve is appreciable, the divider should be inserted in the anode circuit of this valve.

Battery operation. If desired, this amplifier may be operated entirely from accumulators. For battery operation, however, it is preferable to use the special 2-V. filament-cathode triodes and pentodes. These valves would, of course, require slightly different values of resistances and bias potentials.

§ 7. BACKGROUND NOISE

The limit of useful amplification is passed when the amplitude of the amplified background noise is sufficient to affect the recording device. Further amplification then merely provides more power for operating the recording instrument, without increasing the sensitivity. The principal sources of background noise are (*a*) random fluctuations of the first-valve anode currents (i.e. "shot" effect and Johnson effect); (*b*) mechanical vibration (i.e. "microphonics"); and (*c*) electrical disturbances and faulty apparatus.

Shot effect and Johnson effect. As the number of stages in an amplifier is increased, a point is eventually reached when a continuous rushing noise is heard in a pair of telephones connected in the output circuit, due to the amplification of the random fluctuations which occur in the anode current of the first valve. Several factors contribute towards the production of these fluctuations. One of these is the so-called shot effect—namely, the statistical fluctuation of the rate of arrival of electrons at the anode. Although the *percentage* fluctuations become less as the average emission current increases, the *absolute* value of the fluctuations increases with the square root of the current. It is therefore advantageous to operate the first valve with as low an anode current as possible. If there are traces of gas contained within the valve, the space charge, and hence the emission current, will be influenced by the presence of positive ions which may be formed. Further, the

positive ions may produce fluctuations of the grid current. For these reasons, in order to avoid ionization of any residual gas, it is usual to operate the first valve with a low anode potential. A third, and important factor is the Johnson effect[4], that is, the random fluctuating voltage of thermal agitation which is set up in any circuit. In this case, the circuit in question consists of the natural leakage resistance, in parallel with the capacity to earth of the first-valve grid system. The Johnson e.m.f. set up in this circuit is amplified by the first, and by succeeding valves of the amplifier. In practice, it is found that the addition of an external grid leak to the first valve always increases the Johnson effect noise, except when the resistance of the leak is so low that the amplification is also reduced. No grid leak is therefore used for the first valve of this kind of amplifier. A grid leak may, however, be employed in the first stage of an amplifier used in conjunction with a Geiger-Muller counter, for the lower amplification necessary is insufficient to make either the Johnson effect or the shot effect troublesome.

Choice of first valve. A valve must be chosen which, when used in the first stage, will produce an amplified pulse large in comparison with the background noise. Furthermore, the valve should be non-microphonic, that is, insensitive to mechanical vibrations. To secure these qualities, it may be necessary to sacrifice high amplification for this stage.

One of the first valves tried, which, as far as the writer is aware, has not been surpassed for this particular purpose by any other valve yet made in this country, is the now almost obsolete "D.E.V." This is a thoriated filament valve, of low grid capacity. Its grid insulation, while sufficient for the purpose, is not exceptionally high. Unfortunately, it is rather a microphonic valve. These valves are now rather difficult to obtain, and many may have to be tried before one which has a low background noise is found. Though the D.E.V. has been used for this purpose in Europe and America, in the latter continent the first stage valve most favoured is the "259 B"; an amplifier using this type of valve has been described by Dunning[5]. This is a screen-grid type of valve, for which a very low noise level is claimed. Another American valve for which a low noise level is also claimed is the "38", which is used in an amplifier described by Johnson and Neitzert[3]. Unfortunately, the writer has had no opportunity of trying either of these valves, and is not therefore in a position to report on their relative merits. It seems, however, that there is a distinct need in this country for the development of a special valve, designed for the particular purpose of particle counting.

Microphonics. Mechanical vibration may cause trouble by its effect upon the electrodes of the valve or the chamber. In either case, noise is produced by the amplification of the small alternating potentials induced on the grid by the relative vibration of the collecting electrode and the high potential chamber electrode, or of the grid and the anode or filament of the valve. The trouble can be minimized, on the one hand, by paying careful attention to the chamber design, using rigid electrodes as far as possible, and on the other, by weighting the first valve, and mounting it on a very resilient support. Pads of rubber sponge are usually placed under the box which screens the first valve and ionization chamber, and also under

other parts of the apparatus if necessary. If very thin foil electrodes are used in the chamber, such as, for example, in a differential chamber[6], it may even be necessary to resort to hermetical sealing of the chamber to keep out sound waves.

Electrical defects. Noise may be produced by electrical pick-up, a matter already discussed in § 5. Bad connexions in the cathode-heating circuit are a fruitful cause of trouble. For this reason, no valve holders are used, and all connexions *without exception* are well soldered. The only form of switch which is really satisfactory in this part of the circuit, is a simple mercury cup and amalgamated copper dipper. Poor insulation of the grid circuits, and defective anode resistances will also give trouble. Reliable mica dielectric condensers should be used for the coupling circuits, and only wire-wound anode resistances employed. Imperfect insulation of the ionization chamber electrodes, even when a guard ring is provided, will produce a noisy background. For the collecting electrode, quartz or amber insulation should be used if possible, and ebonite should be avoided. The insulation of the high potential electrode should also be of good quality, though quartz is not essential unless potentials of the order of 1000 V. are used.

§ 8. BETA AND GAMMA-RAY DISTURBANCES

Very often, it is necessary to count α-particles or protons in the presence of β and γ-rays. These unwanted radiations produce a background of ionization in the chamber, the magnitude of which fluctuates from moment to moment. If the ionization is very intense, the fluctuations may be so large that it becomes practically impossible to detect the α-particles or protons. The trouble is due to the superposition of a number of small pulses caused by the individual β-particle tracks in the chamber.

Apart from the obvious methods of reducing the effect of γ rays within the chamber itself there are two methods of attacking this problem. In the method developed by Ortner and Stetter[2], an electrometer valve with a highly insulated grid is used in the first stage, and distortionless amplification is provided by a battery-coupled amplifier. The final anode current is then a linear function of the first-valve grid potential, within limits. Ions are collected as quickly as possible by the use of a high field in the chamber. In an oscillograph record of the final anode current taken when protons enter the chamber in the presence of γ-rays, the base line of the trace gradually rises as the first grid charges up under the influence of the β-ray ionization; the movement, of course, follows the random fluctuations in the ionization. When, however, a proton enters the chamber, the trace shows a well-marked jump, or step, which stands out clearly against the rise and the fluctuations. As the grid of the first valve is well insulated, it is necessary to discharge it periodically, to clear it of accumulated charges.

A different method has been employed by Wynn-Williams and Ward[7]. In principle, the method consists in clearing the grid of accumulated charge as quickly as possible, so as to reduce the probability of the superposition of charges caused

by several individual β-particles. By this means, the amplitude of the background fluctuation is kept low. If, however, an α-particle, or a proton, enters the chamber, the sudden collection of a large number of ions produces a pulse which stands out above the background. Unlike that of Ortner and Stetter, this method may be used for continuous high-speed recording. Further, it does not require the use of a battery-coupled amplifier, resistance-capacity coupling of the form already described being quite suitable.

When an oscillograph is used for recording, it is essential that it should have a high natural frequency, to respond quickly to the current variations, otherwise there will be *mechanical* superposition of the individual deflections. As in Ortner and Stetter's method, a high collecting field should be used in the chamber to ensure quick collection of ions.

In Wynn-Williams and Ward's method, owing to the increased Johnson-effect noise which would be produced, it is not possible to effect quick clearing of the grid and so reduce the γ-ray disturbance by the obvious method of connecting a low-resistance leak between the first grid and the cathode. Instead, the same final result is achieved by making the time constant of one stage, usually the second, very short. For example, in figure 3, a small condenser of capacity 0·006 μF. (or less if desired) is connected in series with the main 0·1 μF. condenser of this grid circuit. When not required, this small condenser is simply short-circuited.

Franz[8], and Duncanson and Miller[9], have also described methods similar to the above, whereby α-particles and protons can be recorded in the presence of considerable γ-radiation, by means of "proportional" counters.

§ 9. AMPLIFIERS FOR GEIGER-MULLER COUNTERS

For an amplifier used in conjunction with a Geiger-Muller counter, one, two or three triodes or screen-grid valves are usually sufficient, the number depending upon the conditions under which the counter is operated. Barasch[10] has, however, described a special circuit arrangement which has certain advantages, in which what is known commercially as a "high-frequency pentode" is employed in the first stage. With amplifiers used in conjunction with Geiger-Muller counters, owing to the fewer stages of amplification required, stability may usually be secured with fewer precautions than are necessary in large amplifiers. Efficient electromagnetic screening of the counter and amplifier is however, advisable, and also, if possible, of the high potential source. The high potential for operating the counter may be conveniently obtained from a simple valve rectifying circuit, which should contain resistance-capacity smoothing circuits, and a voltage stabilizing device, such as, for example, that described by Street and Johnson[11].

Figure 4*a* shows the usual form of connexion between the first stage, supposedly a triode, and the tube counter. In figure 4*b*, an alternative arrangement is shown, which has the great advantage that the tube of the counter is at earth potential, and therefore serves as a screen. The negative pulse developed at the wire when the tube counter operates and a current passes through the resistance Z, is applied to

the valve grid through the small air condenser k. Resistance-capacity coupling is, in fact, used between the counter and the valve.

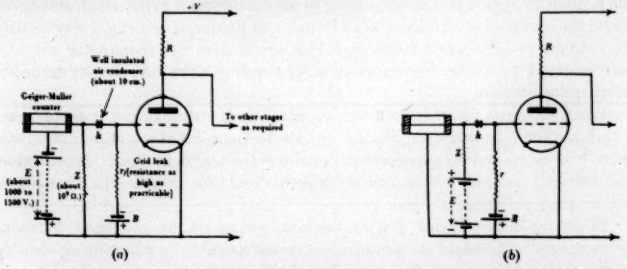


Fig. 4. Alternative connexions for tube counter and first valve (screening arrangements are omitted). The counter and condenser k should be provided with earthed guard rings.

§ 10. COINCIDENCE COUNTING

Occasions occur, for example in cosmic ray experiments, when it is necessary to operate a recording device, or to trip an expansion chamber, only when pulses occur simultaneously in two or more Geiger-Muller counters, each of which is provided with at least one stage of amplification. The amplified pulses from the separate counters are passed into a coincidence circuit. This may, for example,

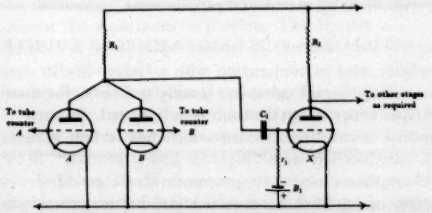


Figure 5. Coincidence counter circuit. Connexions of the counters to valves A and B are as shown in figure 4 a or 4 b.

depend for its operation upon the simultaneous application of separate pulses to both grids of a double grid valve[12]. The more usual method, however, is that due to Rossi[13]. Suppose (figure 5), that valves A and B share a common anode resistance R_1. When both counters operate simultaneously, the grids of both valves become strongly negative, and the two anode currents, combined in R_1, fall to negligible values. A large potential pulse is therefore developed at the junction of the two anodes and the resistance, and is communicated to the grid of valve X by condenser C_1, for further amplification. A coincidence is therefore registered by the thyratron counter connected to the output valve of the system (not shown). If,

however, only one Geiger-Muller counter operates, one valve remains unaffected, and very little change occurs in the value of the current passing through R_1. The resulting potential pulse passed on to valve X, and subsequently amplified, is therefore very small, and it can be arranged that the thyratron counter will not respond to it. The common resistance R_1 can be shared by more than two valves, thus enabling multiple coincidences to be observed.

§ 11. THE RECORDING OF PULSES

The amplified pulses may be recorded (*a*) by ear, using telephones, (*b*) photographically, by means of an oscillograph or (*c*) electromechanically, by means of automatic thyratron counters. Method (*a*) is only suitable for rough counting of small numbers; (*b*) is most useful for giving knowledge of the relative numbers of particles of different ionizing powers, especially when there is much background noise. Method (*c*) has the great advantage of possessing high counting speed, of the order of several thousand particles per minute, without requiring the development and examination of films.

Oscillograph recording. Normally, anode current of the last valve is passed through a current oscillograph. As explained, a high inductance winding may be used to give sufficient sensitivity, if a pentode valve is used in the last stage of the amplifier. In the absence of much background noise, the natural frequency of the oscillograph may be low, with a gain in sensitivity. If, however, there is much β and γ-ray disturbance, a high-frequency instrument is essential. A description has been published[7] of a very simple moving-iron type of oscillograph, which is easily constructed in the laboratory workshop, and answers the purpose very well; in the same paper a brief description is given of the photographic arrangements used.

Whatever type of oscillograph is used, the whole system, amplifier and oscillograph, should always be tested for linearity of response to pulses of various amplitudes, by inducing potential pulses of various magnitudes on the grid of the first valve. The deflection scale can be standardized in ions per millimetre of deflection on the film, by means of a shallow ionization chamber of known depth, through which a narrow beam of α-particles is made to pass. It is usually possible under good conditions to attain a sensitivity of the order of 300 ions per millimetre of deflection. It does not necessarily follow, however, that the sudden production of as few as 300 ions will be detected; sufficient ions must be produced to give a deflection which will stand out clearly against the background. The minimum number of ions detectable, therefore, depends very much upon the experimental conditions.

Some examples of oscillograms taken by Prof. Chadwick have already been published in volume 1 of these *Reports*, p. 284, plate I (1934).

Automatic counting by means of thyratrons. The magnitude and duration of the amplified pulses are usually too small for convenient operation of electro-mechanical counting devices. By means of hot cathode thyratron valves, however, almost unlimited power amplification may be obtained, enabling robust, quick acting counting devices to be operated with ease.

§ 12. CHARACTERISTIC PROPERTIES OF A THYRATRON

Before discussing thyratron counting circuits, a brief account of the general properties of thyratrons will be helpful. The thyratron consists essentially of a triode containing an easily ionizable gas such as mercury vapour or argon, at low pressure. Assuming initially that the grid is at a suitable negative potential, and the anode at a positive potential, only a very small electron current passes between the anode and the cathode; in this respect the thyratron behaves as an ordinary triode. If, however, the potential of the grid is gradually made less negative, a critical potential is eventually reached when the enclosed gas becomes ionized, and a continuous discharge, or luminous arc, is maintained in the gas. As the positive ions from the arc neutralize the space charge of electrons round the cathode, a large anode current may be passed by the thyratron, even at low anode potentials, provided the emitting surface of the cathode is sufficiently great. Quite a small thyratron may pass a current of the order of an ampere. Once the arc has struck, the magnitude of the anode (or arc) current is governed almost entirely by the impedance or resistance of the external anode circuit. The anode potential therefore falls sharply from its previous high value of, say, +200 V., to a low one of the order of +15 V., which is just sufficient to maintain the arc in the thyratron. Further, a sheath of positive ions forms round the grid electrode and prevents any further control of the thyratron current by variation of the grid potential, unless a very high negative potential is applied. Thus, the continuation of the arc current after even a momentary reduction of the negative grid potential, say by the application of a quick positive pulse to the grid, can be regarded as a memory property of the thyratron.

The arc current can be stopped, and the grid allowed to regain control, either by interrupting the anode circuit momentarily, or by causing the anode potential to fall below the minimum arc potential for a time long enough for recombination and diffusion of the ions to take place. This time is of the order of 10^{-4} sec. For a fuller account of the properties and applications of thyratrons, other papers should be consulted[14].

§ 13. SIMPLE THYRATRON COUNTER

The simplest method of using a thyratron for counting is that shown in figure 6*a*. Positive impulses are applied to the grid from the secondary winding of the transformer, through the primary of which the amplifier output-valve anode current passes. (Alternatively, figure 6*b*, a resistance or other impedance may be inserted in the output-valve anode circuit, and the potential pulses developed at the valve anode may be applied through a condenser to the thyratron grid. This method is suitable only if the output-valve anode is made more *positive* when a particle enters the chamber*.) Normally, the thyratron grid is kept well negative by the bias

* In either form of circuit, a.-c. may be used for cathode heating, whether directly or indirectly heated thyratrons are used. Note that direct connexion between the amplifier (which is earthed) and the d.-c. mains, is prevented in (*b*) by the blocking condenser.

battery B, through the resistance r. Assuming that a positive pulse arrives sufficiently great to overcome this excess bias, the thyratron arc is struck, and current passes through the anode circuit. The armature A of the counting meter is attracted by the electromagnet M, and in moving first operates the numerical dial D, increasing the number registered by unity, and finally opens the anode circuit at contact K. The anode current ceases, the grid regains control, and the armature and contacts revert to their normal positions, ready for the next impulse. Discrimination between large and small voltage impulses can be effected by variation of the opposing grid bias, or by means of the potential divider r, which determines what fraction of the impulse is applied to the grid. A resistance, in series with a condenser, is connected across contact K, to prevent arcing.

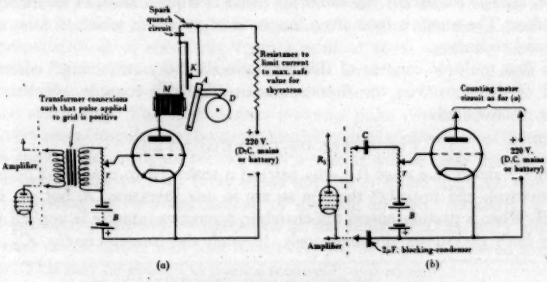


Figure 6. Alternative single thyratron counting circuits.

§ 14. RESOLVING TIME

The efficiency of a simple counter such as the one just described obviously depends upon the operating time of the counting device. Very often, telephone exchange "registers" with increased spring tensioning are used for this purpose. With these, the resolving time can be pushed down to about $\frac{1}{20}$, or even $\frac{1}{25}$ sec. If, however, an automatic telephone "uniselector" rotary switch be used, the resolving time can be reduced to $\frac{1}{80}$, or $\frac{1}{120}$ sec. It is not generally realized, however, what a large fraction of the particles can be missed at even low counting speeds by such a simple counter, owing to the chance arrival of particles during the time when the meter is in the process of recording. For a resolving time of $\frac{1}{20}$ sec. the probable fraction missed when the rates of entry of particles into the chamber are 30, 60, 120 and 240 per min. are respectively 2·4, 4·8, 9·1 and 16·6 per cent. With resolving times of $\frac{1}{80}$ or $\frac{1}{120}$ sec., these fractions are proportionately less. This emphasizes the need for a counter of very low resolving time, if fast counting is contemplated. Compared with this, the "scale-of-two" counter shown in figure 8, § 16, has a resolving time of about $\frac{1}{4000}$ to $\frac{1}{8000}$ sec. Assuming a value of $\frac{1}{5000}$ sec. and that particles arrive at the rate of 3000 per min., the probable fraction missed is only

1 per cent. Unless therefore the number of particles counted in an observation is greater than 10,000, the error involved at counting speeds up to 3000 per min. is less than the statistical fluctuation.

Multiple thyratron counters. The limitation of the counting speed in the circuit just described is due entirely to the mechanical inertia of the meter. A thyratron may be tripped in a few microseconds, and reset, or de-ionized, in a time of the order of 10^{-4} to 10^{-3} sec. Special multiple thyratron circuits were therefore devised by the writer to obviate the limitation of counting speed imposed by the mechanical counting device. In these circuits it is arranged that while one thyratron is actually responding to an impulse, it is at the same time preparing another thyratron to respond to the next impulse. After a certain definite number of impulses have been counted, *entirely electrically*, the counting meter is tripped once, to record a group of impulses. The meter is thus given longer time periods in which to carry out its mechanical recording.

The first multiple counter of this kind was the "thyratron ring" counter[15]. Instead of this, however, the thyratron scale-of-two[16] counter, which is more efficient, is now used.

§ 15. SCALE-OF-TWO COUNTER

Figure 7 shows one stage (i.e. one pair) of a scale-of-two counter. The circuit is symmetrical, and initially* there is an arc in one thyratron, *A*, but not in the other, *B*. When a particle enters the chamber, a positive impulse is applied to the grids of both thyratrons simultaneously, through the condensers C_A, C_B. That

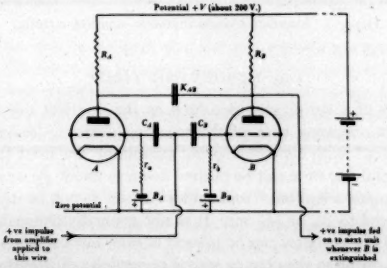


Figure 7. Single unit of thyratron "scale-of-two" counter.

applied to *A* has no effect, for there is already an arc in *A*. That applied to *B*, however, causes the arc to strike in this thyratron. The anode potential of *B* now drops sharply, from a value of, say, +200 V., to the arc voltage of about +15 V. A large negative potential pulse of about −185 V. is therefore communicated from the anode of *B* to the anode of *A* through the "transfer" condenser K_{AB}. The potential of the anode of *A*, which, owing to the presence of the arc in this thyratron,

* When first switched on, anode potential is applied to one thyratron before the other.

was previously about +15 V., therefore drops suddenly by 185 V. to about −170 V. If the values of R and K are not too small, this negative anode potential will persist long enough for the arc in A to be extinguished, and de-ionization to take place, in which case the grid of A regains control before the full anode potential is restored. Thyratron A will therefore be "reset" [i.e. ready to respond to the next impulse], while the arc in B continues. As the circuit is symmetrical, it is clear that successive impulses "transfer" the position of the arc in this pair from A to B or vice versa. The resolving time is dependent upon the rapidity of de-ionization, and quite independent of mechanical inertia.

Additional stages. The change of anode potential of, say, B, is different when the arc in B strikes, from when it is extinguished. On *striking*, as explained, the anode potential drops sharply by about 185 V. When the next impulse arrives, however, and B is *reset* by the striking of A there is first of all a sharp drop in the anode potential of B, of about 185 V., to a value of about −170 V., but this is immediately followed by a *rapid exponential rise* from say −170 V., right up to the normal value of say +200 V., that is, a rise of about 370 V.; this rise is double the initial sharp drop. This large *positive* potential pulse developed at the anode of B at every *second* impulse may be applied to the grids of another pair of thyratrons, C, D, as shown later in figure 8*b*. In such a case, it will be clear that the arc in stage CD will change over for every *second* impulse to which stage AB responds, thereby providing an electrical method of dividing by two. Similarly, a large positive potential pulse will be developed at the anode of D at every alternate change-over of the arc in CD, that is, at every *fourth* impulse received from the amplifier. This may be applied to a third stage EF, and so on.

Counting scale. As explained later, three stages are usually necessary; with such an arrangement there are 2^3, or 8, different arc positions. As impulses arrive, the positions of the arcs go through the cycle:

ACE,	*BCE*,	*ADE*,	*BDE*,	*ACF*,	*BCF*,	*ADF*,	*BDF*,	*ACE*, etc.
(0)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	

When, at the eighth impulse, the system returns to the original configuration ACE, a large positive pulse is developed at the anode of F. This may be applied to a simple thyratron counting circuit similar to that of figure 6 (*b*), operating a mechanical meter. If the number of impulses counted during a run is not a multiple of eight, this will be indicated by the final positions of the arcs. For example, if the meter has been operated nine times, and the arcs are left in thyratrons B, C and F, the number of impulses counted is $(9\times 8)+5$, or 77. In practice, instead of referring to the above table of arc configurations, values of 1, 2 and 4 (i.e. 2^0, 2^1 and 2^2) are assigned to thyratrons B, D and F; if arcs are left in any of these thyratrons, the appropriate numbers are merely added to eight times the meter reading. This conversion from the scale-of-two notation to the ordinary decimal scale may be done electrically, as explained in § 17.

Number of stages needed. Any number of stages may be employed in a scale-of-two counter. It is only necessary, however, to provide a number N of stages

sufficient to ensure that it is less probable that a group of 2^N impulses will arrive in the meter recording time, than a pair of impulses in the resolving time of the first stage of the counter. If the fullest advantages of the short resolving time of the first pair are to be realized, it is usually necessary to use three stages.

Use of valves in scale-of-two counters. A novel type of counter has recently been used by Lewis. This makes use of the scale-of-two principle, but instead of thyratrons employs triode valves in the counting units, with a special form of coupling between the stages. An account of this type of counter is to be published shortly.

§ 16. PRACTICAL DETAILS OF SCALE-OF-TWO COUNTER

Figure 8 shows the circuit arrangement used at the Cavendish Laboratory. In figure 8*b*, the three stages, *AB*, *CD* and *EF*, and the metering thyratron, *G*, will be recognized. Note that the links between thyratron *D* and stage *EF*, and between thyratron *F* and stage *G*, consist of 20,000-ohm resistances connected, not to the anodes of *D* and *F*, but to the middle of the anode resistances. This reduces the tendency for a pulse to be fed back from one stage to the previous one. Note also that in the link between thyratron *B* and stage *CD*, a "trap" valve is inserted; this is a simple single stage valve amplifier, with a 1 : 1 ratio auto-transformer in its anode circuit to avoid reversal of the pulse polarity. Amplified pulses can therefore pass forwards, but not backwards, through this link; this extra precaution is taken because thyratrons *A* and *B* are operated with less margin of grid bias than the others.

Suppressing-condensers. In the first stage will be seen an extra pair of small condensers, called suppressing-condensers. The object of these is to prevent jamming (i.e. the simultaneous striking of arcs in *A* and *B*), which occurs if a positive impulse arrives at the grid before the anode potential of an extinguished thyratron has had time to recover to its full value. These condensers transmit a large negative pulse to the grid of the extinguished thyratron, and until this pulse has died away (more slowly than the anode potential recovers) the arc cannot be restruck in that thyratron. The addition of these condensers also enables the resolving time to be reduced from about $\frac{1}{1500}$ sec., to about $\frac{1}{4000}$ to $\frac{1}{8000}$ sec. The matter is dealt with more fully in a paper by Lewis[17].

Clearing. In order to start from the zero position (*ACE*), the multiple clearing contact is operated, either by hand or by a clearing relay. This contact "earths", in correct order, the grids of thyratrons *A*, *C* and *E*, thus restoring the arcs from whatever positions they previously occupied to the normal one. Other contacts (not shown), operated by the clearing key or relay, cause the mechanical meter or other registering device to run rapidly to the 0000 position, ready for the next count.

Discriminator circuit. The impulses coming from the amplifier will, in general, be of various sizes; it is usually desired to record only those greater than a certain size. This might be done, as in the case of the simple counter previously described, by adjustment of the grid bias of thyratrons *A* and *B*. It is preferable, however,

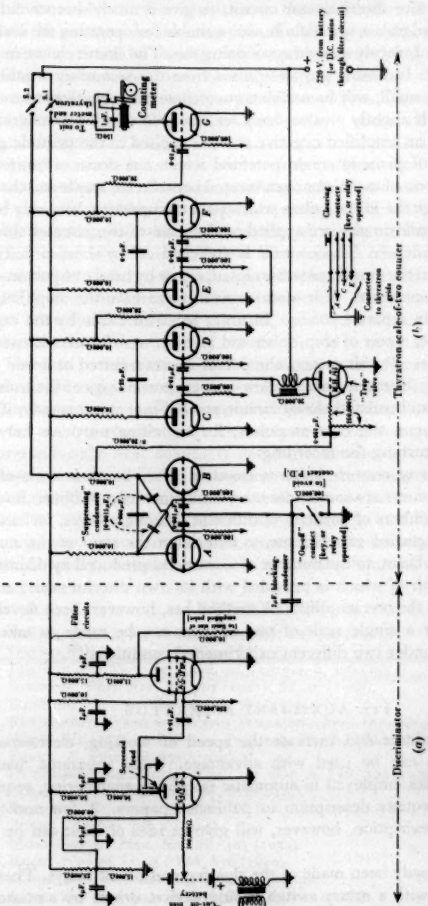


Figure 8. Detailed diagram of complete scale-of-two counting circuit. For simplicity, grid-bias batteries and cathode-heating arrangements are not shown. All valves: indirectly heated from 4-V. a.c. All thyratrons: directly heated from 2-V. a.c. Type of thyratron: "Mazda BT 8" for *A*, *B*, *C*, *D*, *E* and *F*, and "Mazda BT 1" for *G*. Bias potentials: "cut-off" up to − 100 V.; A.C./Pen. − 15 V.; M.H./41, − 6 V. Thyratrons *A* and *B*, − 6 to − 9 V.; *C*, *D*, *E* and *F*, − 30 V.; *G*, − 15 V. All "earths" to be interpreted as common return lead; no part of this circuit is actually connected to an earth wire.

to use a separate valve discriminator circuit, to give a much sharper cut-off, and to amplify the passed pulses to a definite size, suitable for operating the scale-of-two counter. There are obviously many ways of doing this. The circuit shown in figure 8*a* is very convenient. Incoming positive pulses from the secondary winding of the transformer, if too small, will be unable to overcome the negative "cut-off" bias battery potential. If slightly greater, however, the grid of the screen-grid valve is made positive, and an amplified negative pulse is applied to the pentode grid. The maximum change of pentode anode potential which can occur is limited by the presence of the 10,000-ohm resistance connected between the anode and the cathode. Thus, however large the initial pulses which pass the opposing bias may be, pulses of approximately uniform size are applied to the scale-of-two counter through the 2 μF. blocking condenser. The counter is switched off by short-circuiting these pulses to earth, when the on-off contact is closed, either by hand or by an on-off relay.

Connexion with amplifier. It is often convenient to have the amplifier and the thyratron counter in separate rooms. In order to avoid losses by the capacity of a long screened line, a pair of step-down and step-up transformers is used, one at each end of the line. By this means, the pulses are transmitted at lower potential and higher current. Further, since the low-resistance windings of the transformers and the lines form an insulated closed circuit, small direct currents may, if desired, be superimposed upon the current pulses, for signalling purposes between the rooms, without disturbing the recording.

Simultaneous use of several thyratron counters. Two or more scale-of-two, or single thyratron counters are sometimes used simultaneously, to obtain information as to the relative numbers of particles of different ionizing powers, for example, in high voltage disintegration experiments, to determine the ratio of the number of protons in the main beam, to the number of α-particles produced by disintegration. These counters, each of which is provided with its own discriminator, are all fed with impulses from the one amplifier. A method has, however, been developed by the writer whereby a single scale-of-two counter can be made to take rapidly alternating counts under two different experimental conditions[18].

§ 17. AUXILIARY APPARATUS

In order to facilitate and increase the speed of working, electromechanical auxiliary apparatus can be used with advantage. Such apparatus, usually incorporating principles employed in automatic telephone engineering, requires too much space for adequate description in published papers. A few non-technical words of general description, however, will give an idea of what can be done in this way.

Mention has already been made of the clearing and on-off relays. These relays are interconnected with a rotary switch timing device, driven by a master clock, which switches on the counter for a definite predetermined period. In taking a run, therefore, all that is required of the observer is that he should first operate the clearing relay by pressing a button, and then operate a start relay by pressing

a second button. No further attention is then needed until the timing device switches off the counter. In the meantime, the observer is able to concentrate upon other matters incidental to the experiment, without continual reference to a stop-watch. Interlocking connexions prevent wrong operation of the device. By such means, for example, as many as fifty separate two-minute observations can be made with a rapidly decaying radium (B+C) source in a period of two hours without difficulty or fatigue.

Further refinements have been made which perform the "translation" of the scale-of-two number into the corresponding scale-of-ten number, and record all details of the run in plain figures on a paper tape. The arithmetic is performed electrically, by using, instead of the telephone call register in the circuit of thyratron *G*, a pair of uniselector rotary switches. The contacts of these switches are cross wired according to a scheme representing multiplication by eight, so that each single step made by the switch when the thyratron strikes "adds" eight to the previous setting. A set of other associated contacts, operated by three relays connected in the anode circuits of thyratrons *B*, *D* and *F*, add 1, 2 or 4 as required to this setting. The final result appears as a four-figure scale-of-ten number on a small lamp dial. At the end of a run, a rotary printing device hunts for, and prints in turn on a paper tape, the four digits representing the thousands, hundreds, tens and units of the number counted, and then operates the clearing relay, thus resetting the apparatus for the next count. If desired, the same device numbers the runs on the tape consecutively, and also prints other experimental data, such as the duration of, and the mean time at which each run was taken. Such apparatus can be arranged to run continuously if necessary, carrying out a long series of experimental runs, say overnight, without attention. Statistical data may thus be accumulated rapidly, without the necessity for long hours of routine counting by an observer.

REFERENCES

(1) Greinacher. *Z. Phys.* **36**, 364 (1926); **44**, 319 (1927).
(2) Ortner and Stetter. *Z. Phys.* **54**, 449 (1929).
(3) Johnson and Neitzert. *Rev. sci. Instrum.* **5**, 196 (1934).
(4) Johnson. *Phys. Rev.* **29**, 367 (1927); **32**, 97 (1928).
(5) Dunning. *Rev. sci. Instrum.* **5**, 387 (1934).
(6) Rutherford, Ward and Wynn-Williams. *Proc. roy. Soc.* A, **129**, 211 (1930).
(7) Wynn-Williams and Ward. *Proc. roy. Soc.* A, **131**, 391 (1931).
(8) Franz. *Z. Phys.* **63**, 370 (1930).
(9) Duncanson and Miller. *Proc. roy. Soc.* A, **146**, 396 (1934).
(10) Barasch. *Proc. phys. Soc.* **47**, 824 (1935).
(11) Street and Johnson. *J. Franklin Inst.* **214**, 155 (1932).
(12) Bothe. *Z. Phys.* **59**, 1 (1930).
Mott-Smith. *Phys. Rev.* **39**, 403 (1932).
(13) Rossi. *Nature*, Lond., **125**, 636 (1930).
(14) Hull. *Gen. elect. Rev.* **32**, 213, 390 (1929).
Lewer and Dunham. *G.E.C. Jl.* **3**, 67, 119 (1932).
(15) Wynn-Williams. *Proc. roy. Soc.* A, **132**, 295 (1931).
(16) Wynn-Williams. *Proc. roy. Soc.* A, **136**, 312 (1932).
(17) Lewis. *Proc. Camb. phil. Soc.* **30**, 543 (1934).
(18) Wynn-Williams. *Proc. phys. Soc.* **46**, 303 (1934).

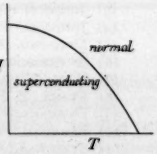
SUPERCONDUCTIVITY AND THE THEORY OF METALS

By A. H. WILSON, Trinity College, Cambridge

§ 1. INTRODUCTION

Our knowledge of the phenomenon of superconductivity has greatly increased during the last few years, but we seem as far as ever from a microscopic theory of the effect. The application of thermodynamics to the problem has, however, had some success in correlating various experimental facts, and it is with this aspect of the theory that this report mainly deals.

An account of the empirical data has already been given in these reports by J. C. McLennan[1], and it will therefore not be necessary to give more than an outline of the phenomena dealt with. The electrical resistance of metals, which at ordinary temperatures is proportional to the absolute temperature, approaches a constant value as the temperature approaches the absolute zero. This constant resistance is known as the residual resistance and is supposed to be due to impurities and irregularities in the metal lattice caused by cracks or strains. At a low temperature (7·2° K. for lead) the residual resistance of certain metals disappears more or less abruptly, and it is probable that the transition to the superconducting state would be absolutely sharp for a perfectly pure substance. If a superconducting metal at a temperature below the transition temperature is placed in a gradually increasing magnetic field, the electrical resistance reappears when a certain critical field has been reached. This field depends on the temperature, being zero at the transition temperature and increasing as the temperature is lowered. The limiting value at the absolute zero is almost certainly finite, and so the ranges of temperature and magnetic field for which a metal is superconducting or normal are as shown in the figure. Since the resistance of a superconductor is zero, very large currents can exist, but it is found that there is an upper limit to the current density. This is supposed to be a secondary effect due to the magnetic field produced by the current, a suggestion due to Silsbee[2], and it has been verified that for pure metals the magnetic field at the surface of a wire carrying a limiting current is just the critical field necessary to destroy superconductivity.



Experiments by Meissner and Ochsenfeld[3] on the magnetic behaviour of superconductors show that the disappearance of the residual resistance is not the only, and perhaps not even the most fundamental, phenomenon which occurs. If a body is in the superconducting state, Faraday's induction law shows that the magnetic induction cannot be changed in any portion of the conductor by any process whatever. If the external magnetic field is changed, persistent surface

currents will be set up so as to neutralize the change in flux. It has, however, been found that not only is the magnetic induction invariable in a superconductor, but that it has the value zero under certain ideal conditions, which are that the body should be a pure metal and be in the form of a long ellipsoid with its greatest axis parallel to the magnetic field. This means that the properties of a superconductor are independent of its previous history, and that the transition from the normal to the superconducting state is reversible. It therefore seems reasonable to regard the transition as a phase change similar to that which occurs on melting, and to see what information thermodynamical arguments can give.

§ 2. THE THERMODYNAMICS OF SUPERCONDUCTIVITY[4]

In this section we shall only deal with ideal superconductors in which the magnetic induction is zero, and we shall assume that the body is in the form of a long rod parallel to the external magnetic field H, which is supposed to be uniform. The field inside the rod will then be H also, and since $B=0$ there will be an induced magnetic moment of amount $-H/4\pi$ per unit volume. If U is the internal energy and S the entropy, both per unit volume, then since the magnetic energy is $(1/4\pi)\int HdB$ per unit volume[5], we can write the first and second laws of thermodynamics in the form

$$\begin{aligned} dU &= TdS+\frac{1}{4\pi}HdB \\ &= TdS+HdI+\frac{1}{4\pi}HdH \qquad \text{.....(1),} \end{aligned}$$

where I is the intensity of magnetization. In a superconductor the sum of the last two terms in equation (1) is zero, but it is convenient to disregard the part $H^2/8\pi$ of the magnetic energy which exists when the body is absent, and to call U' the internal energy, where

$$U'=U-H^2/8\pi.$$

This means that the magnetic energy of a normal metal is zero, provided we neglect its very small susceptibility, while the magnetic energy of a body in the superconducting state is HdI. This change of energy zero is unimportant, since the only essential point is that dU' is a perfect differential, where

$$dU'=TdS+HdI \qquad \text{.....(1 a).}$$

Now I is independent of the temperature and is actually $-H/4\pi$ for a superconductor (zero for a metal in its normal state), so that there are no terms involving both H and T. The specific heat is therefore independent of H. The formal proof is as follows. We have, taking T and H as the independent variables,

$$TdS=\frac{\partial U'}{\partial T}dT+\left(\frac{\partial U'}{\partial H}-\frac{\partial I}{\partial H}\right)dH.$$

Hence, since dS is a perfect differential, we find

$$\frac{1}{T^2}\frac{\partial U'}{\partial H}-\frac{1}{T^2}\frac{\partial I}{\partial H}=0,$$

and so $$TdS=\frac{\partial U'}{\partial T}\,dT=cdT \qquad \text{.....(2),}$$

where c is the specific heat per unit volume.

The function G, defined by

$$G=U'-ST-HI \qquad \text{.....(3),}$$

plays the part of the free energy. From equation (1 a) we have

$$dG=-SdT-IdH \qquad \text{.....(4),}$$

and this can readily be integrated so as to give G explicitly, since from equation (2)

$$S=\int^T \frac{c}{T}\,dT \qquad \text{.....(5),}$$

and I is independent of T. An integration by parts gives

$$G=\int^T cdT-T\int^T \frac{c}{T}\,dT+\frac{H^2}{8\pi} \qquad \text{.....(6).}$$

In passing from the superconducting to the normal state there is an amount of heat Q absorbed per unit volume, and this is related to the temperature and field at the transition point by a relation similar to Clapeyron's equation in the thermodynamics of fluids. Denoting the superconducting and normal phases by suffixes s and n, the relation

$$G_s=G_n$$

must hold all along the transition curve separating the two phases in the H-T diagram shown in the figure. For we have

$$\left(\frac{\partial G}{\partial T}\right)_H=-S, \quad \left(\frac{\partial G}{\partial H}\right)_T=-I \qquad \text{.....(7),}$$

and, if G_s and G_n were not equal, there would be an infinite discontinuity in S and I on crossing the boundary between the phases. We must therefore have

$$dG_s=dG_n,$$

where dG denotes a variation along the phase boundary curve, or from equation (4)

$$\frac{dH}{dT}=-\frac{S_s-S_n}{I_s-I_n} \qquad \text{.....(8).}$$

Now if we neglect the very small magnetic susceptibility of the normal metal, the equations (1 a) to (6) hold for the normal metal also, with $I=0$ and therefore with $H=0$ in equation (6). Also

$$T(S_n-S_s)=Q, \quad I_s=-H/4\pi,$$

and equation (8) takes the form

$$\frac{dH}{dT}=-\frac{4\pi Q}{HT} \qquad \text{.....(9),}$$

which gives Q in terms of the slope of the phase boundary curve. This equation obviously fails when $H=0$, that is at the ordinary transition temperature T_0 in zero field. Experimentally dH/dT is not infinite at this point, and so Q must be zero. This is verified experimentally, there being no heat of transition for $T=T_0$ but only a discontinuity in the specific heat.

We can obtain further relations, which are often more useful than equation (9), in the following way. Since $G_s = G_n$, and since $H = 0$ on the boundary curve when $T = T_0$, we have all along the boundary curve

$$\frac{H^2}{8\pi} = \int_T^{T_0} (c_s - c_n)\, dT - T \int_T^{T_0} \frac{c_s - c_n}{T}\, dT \qquad \text{.....(10).}$$

This enables us to calculate the magnetic threshold if the difference of the specific heats is known as a function of the temperature. Equation (10) can be transformed into another useful form by differentiating twice with respect to T, obtaining

$$\left(\frac{dH}{dT}\right)^2 + H\frac{d^2H}{dT^2} = 4\pi \frac{c_s - c_n}{T} \qquad \text{.....(11).}$$

This is the most convenient form near $T = T_0$, and if we put $H = 0$ we obtain Rutgers' formula[6]

$$\left(\frac{dH}{dT}\right)^2_{T=T_0} = 4\pi \frac{c_s - c_n}{T_0} \qquad \text{.....(12),}$$

which was first derived by considering the equilibrium to be of the type which Ehrenfest called "of the second kind".

We can obtain an expression for Q from the function

$$U' - HI = G + ST = \int^T c dT - \tfrac{1}{2} IH \qquad \text{.....(13).}$$

For, using equation (1 *a*),

$$d(U' - HI) = TdS - IdH \qquad \text{.....(14),}$$

and so the heat absorbed when H is constant is the change in $(U' - HI)$. Since the transition to the normal state takes place at constant H and not constant I we have

$$\begin{aligned} Q &= (U' - HI)_n - (U' - HI)_s \\ &= \int_T^{T_0} (c_s - c_n)\, dT - \frac{H^2}{8\pi} \qquad \text{.....(15).} \end{aligned}$$

The equations (10) and (15) can be combined so as to reproduce equation (9).

§ 3. NUMERICAL RESULTS

The validity of the assumption that the transition from the normal to the superconducting state is a reversible one can be tested by comparing the experimental results with the theoretical relations (9), (10), (12) and (15). It is found that, when the transition takes in zero field, there is no heat of transition and that for tin and thallium the discontinuity in the specific heat is in good agreement with Rutgers' relation (12). The following figures are given by Keesom and Kok[7].

Table 1. The discontinuity in the specific heat in zero field ($\Delta c = c_s - c_n$)

	T_0 (° K.)	dH/dT (G./° K.)	Δc (cal./° K. per mol)	
			Calculated	Observed
Tin	3·71	151·2	0·00229	0·0024
Thallium	2·36	137·4	0·00144	0·00148

The results available for testing equation (9) are very meagre, but the relation seems well verified for thallium. For example, when $H=33{\cdot}6$ gauss the observed heat of transition is 0·000276 calorie per mol, while the calculated value is 0·000295.

Some further interesting numerical relations have been given by Kok[8]. The phase boundary curve in the H-T plane is a parabola for certain metals, in particular for indium, thallium and tin, and therefore has the equation

$$H=H_m\,(1-T^2/T_0^2) \qquad \text{.....(16).}$$

Equation (11) shows that (c_s-c_n) is then completely determined by the two constants H_m and T_0, the exact relation being

$$c_s-c_n=-\frac{H_m^2\,T}{2\pi T_0^2}+\frac{3H_m^2\,T^3}{2\pi T_0^4} \qquad \text{.....(17).}$$

Now the specific heat of a normal metal consists of two parts, one being due to the thermal vibrations of the lattice, while the other contribution comes from the free electrons. The former is of amount αT^3, where α is a constant, and the latter is γT per unit volume[9], where

$$\gamma=\frac{4\pi^3\,mk^2}{3h^2}\left(\frac{3n}{\pi}\right)^{\frac{1}{3}} \qquad \text{.....(18).}$$

Here n is the number of free electrons per unit volume. This formula is derived on the assumption that the electrons obey the Fermi-Dirac statistics and that they move under no forces in the metal. It is possible to take crystal forces into account in various ways, the simplest being to ascribe formally an increased effective mass m^* to the electrons. The electronic specific heat is extremely small and only becomes important at low temperatures. Recently Keesom and Kok[10] have found that the specific heat of silver below 3° K. is proportional to T and has almost exactly the value given by equation (18) if n is put equal to the number of atoms per c.c., that is if there is one free electron per atom. Now it seems probable from the experiments that in the superconducting state the total specific heat is proportional to T^3, and if this is so we have, on combining equations (17) and (18), the relation

$$\gamma=\frac{H_m^2}{2\pi T_0^2} \qquad \text{.....(19).}$$

The figures given in table 2 show that the agreement is surprisingly good, if n is taken to be the number of valency electrons per unit volume.

Table 2

	T_0 (° K.)	H_m (G.)	$n\times10^{-22}$	$\gamma\times10^4$ (cal./(° K.)^2 per mol)	
				Calculated from equation (19)	Calculated from equation (18)
Thallium	2·36	163	10·8	3·06	3·10
Tin	3·71	281	17·1	3·08	3·03
Indium	3·37	237	11·5	2·96	2·97

Further, when $T = T_0$ we have

$$c_s - c_n = 2\gamma T_0 \qquad \text{.....(20),}$$

and this will be valid for those substances for which equations (12) and (19) are true.

§ 4. THE INFLUENCE OF SHAPE AND PURITY

The thermodynamic theory based on the assumption that $B = 0$ seems to be well substantiated by the experiments on pure substances in the form of rods, provided the magnetic field is longitudinal. With such an arrangement, the magnetic field H is practically uniform everywhere, and if a superconductor is placed in a gradually increasing field nothing happens until the threshold value is reached, when the whole body passes over reversibly into the normal state. There is, therefore, never more than one phase present in the body at any time, and it is only when this state of affairs is realized that the theory is valid in its entirety. Let us consider a long circular superconducting cylinder in a uniform magnetic field H which is perpendicular to the axis of the cylinder. Assuming that $B = 0$, the field outside the body varies from zero at the ends of the diameter parallel to the field to a maximum value of $2H$ at the ends of the diameter perpendicular to the field. Hence the field at the surface of the superconductor will have reached the threshold value when the applied field is only half that value. A slight increase in the field will not, however, cause the whole of the body to pass over into the normal state, for then the internal field would be only half the threshold value. Experiments by de Haas and Casimir-Jonker[11] showed that with increasing field the magnetic induction gradually penetrated the cylinder, but their experimental arrangement was not such as to give conclusive evidence as to the exact manner of the penetration. It has also been found[12] that the restoration of the resistance is gradual and follows a similar course.

The situation has to a certain extent been cleared up by more recent experiments on the behaviour of ellipsoids, and in particular of spheres[13], in a uniform field H_1. If $4\pi\alpha$ is the demagnetizing factor ($\alpha = \frac{1}{3}$ for a sphere), the internal field H is uniform and equal to $H_1 - 4\pi\alpha I$. Since $B = H + 4\pi I$, we must have

$$H_1 = H(1-\alpha) + \alpha B.$$

Now, if $B = 0$ as in a superconductor, the internal field will have reached the critical value H_c when $H_1 = H_c(1-\alpha)$. It is found that, when H_1 exceeds this value, the internal field remains equal to H_c but that the induction penetrates the ellipsoid uniformly and has the value $H_1/\alpha + H_c(1 - 1/\alpha)$. The ellipsoid is then neither in the superconducting nor in the normal state, but in what may be called an intermediate state. When H_1 reaches the value H_c the body is in the normal state, and for higher fields $H = B = H_1$. The nature of the intermediate state is still obscure, and all that can be said at present is that there is a resistance which increases with the induction.

A somewhat similar effect is found[14] when a superconducting ring of thallium, in which persistent currents are flowing, has its temperature gradually increased. The threshold in the absence of a field is 2·21° K., but when there are persistent

currents due to the removal of 26·4 gauss a change starts to take place at 2·17° K. This change is a gradual diminution in the persistent current as shown by a reduction in the magnetic field of the coil. This temperature of 2·17° K. is just the transition temperature corresponding to the magnetic field at the surface of the ring produced by the persistent current. It would therefore seem that the disappearance of the persistent current is not an irreversible effect caused by the resistance coming into being, but is primarily due to the entry of the magnetic induction into the metal. It must, however, be admitted that none of these effects is clearly understood at the moment.

The phenomena observed with impure metals, and in particular with alloys[15], are still more puzzling. In the first place the magnetic induction does not become zero when an alloy in a magnetic field is made superconducting either by lowering the temperature or reducing the field. Instead a portion of the flux becomes locked in, and in general the percentage locked in depends on the purity of the specimen, being greater the more irregular the alloy. It has been suggested by Mendelssohn and others that this behaviour might be explained by assuming that different portions of an alloy may have different threshold fields, the threshold depending on the purity. It would then be possible to have a superconducting body with non-superconducting inclusions, and any flux in the non-superconducting regions would be permanently locked in so long as the rest of the body remained superconducting.

The disturbance of superconductivity in an alloy by a magnetic field shows many anomalous features. In general the threshold field for alloys is very high, if by the threshold we understand the field which restores the resistance. On the other hand the field which will cause the induction to start penetrating the alloy is not equal to the threshold for resistance, but is many times smaller. Further, when the interruption of superconductivity is caused by increasing the current density, the surface field is very much less than the threshold for an external field, and it is not equal to the threshold for penetration, so that Silsbee's hypothesis is no longer true. The thermal properties are just as anomalous. Since the threshold fields are high for alloys there should, according to Rutgers' equation (12), be a large discontinuity in the specific heat. In fact there is no discontinuity of anything like the amount predicted, but there is the further difficulty that there are three possible definitions of the threshold field, which would give widely differing results, and it is not easy to see which of them, if any, should be connected with the specific heat.

§ 5. OTHER PROPERTIES OF SUPERCONDUCTORS

It has been shown by de Haas and Bremmer[16] that the thermal conductivity of a metal is altered when it becomes superconducting. The thermal resistance is constant at room temperatures and falls as the temperature is lowered. At helium temperatures it has a minimum and for very low temperatures it rises again. There is no discontinuity at the transition temperature for a superconductor. de Haas and Bremmer found that when superconductivity is destroyed by a magnetic field the thermal resistance is decreased. The temperatures at which they were working

were below the temperature of the minimum in the resistance curve. This result is not very easy to interpret since little is known about the theory of heat conduction at low temperatures. Heat can be transported in two ways in a metal—by the lattice vibrations, as in an insulator, and by the free electrons. These two mechanisms are probably not independent, but considering only conduction by the free electrons we have to take into account both the impurities and the heat vibrations of the crystal lattice in producing the resistance. If we can assume that these two effects are independent the total resistance will be the sum of the two separate resistances. Now it can be shown[17] that the thermal resistance caused by impurities is inversely proportional to the absolute temperature, so that if κ is the thermal conductivity

$$\frac{1}{\kappa}=\frac{1}{\kappa_0 T}+\frac{1}{\kappa_i} \qquad \text{.....(21),}$$

where κ_0 is a constant, and κ_i is the thermal conductivity of an ideal crystal devoid of impurities. At low temperatures κ_i is proportional to T^{-2}, while at high temperatures it is constant. This equation, if true, shows that $1/\kappa$ reaches a minimum value and then increases as T decreases. The temperature of the minimum will be lower the purer the crystal.

Since the destruction of superconductivity by a magnetic field increases the thermal conductivity it seems necessary to suppose that in the superconducting state a proportion of the electrons is not available for carrying the heat current. This hypothesis receives some support from the fact that the specific heat of a superconductor obeys a T^3 law, which shows that the valency electrons do not have the properties of a degenerate gas obeying the Fermi-Dirac statistics. Further evidence for this is that there are no thermoelectric effects between superconductors[18]. On the other hand the derivation of equation (21) is extremely doubtful, since it depends on the existence of a free path, and it is known[17] that a free path cannot be defined at low temperatures. This means that the various resistances are not additive, and the lowering of the thermal resistance by a magnetic field may be due to some secondary effect not worked out at present.

§ 6. THEORIES OF SUPERCONDUCTIVITY

No headway at all has been made in understanding the mechanism of superconductivity. The most promising line of attack is a purely formal theory by F. and H. London[19], which attempts to explain some of the properties of the superconducting phase, without troubling about the reasons for its existence. The existence of a transition temperature and of a threshold magnetic field lie quite out of the domain of any theory which has so far been proposed.

The Maxwell equations of the electromagnetic field are indeterminate unless the constitutive equations are also known. Since Ohm's law does not hold in a superconductor, we must have some other relation connecting the current density $\mathbf{J}$ with the field strength. The most obvious one is the classical relation

$$\frac{d\mathbf{J}}{dt}=\frac{ne^2}{m}\mathbf{E} \qquad \text{.....(22),}$$

where n is the number of free electrons per unit volume and $\mathbf{E}$ is the electric field. This assumes that the electrons are accelerated by a field according to the laws of classical mechanics. It probably gives correctly any changes which occur provided the body originally became superconducting in the absence of any field, but it cannot account for the expulsion of the magnetic induction. We should therefore have to postulate in addition $\mathbf{B}=0$.

The Londons' theory discards equation (22) and replaces it by a condition which ensures that no magnetic field effectively enters a superconductor. According to them the permeability is unity, so that the magnetic field and induction are the same, and the single constitutive relation is

$$\text{curl } \mathbf{J} = -\frac{n\epsilon^2}{mc}\mathbf{H} \qquad \text{.....(23).}$$

Combining this with the equations

$$\text{curl } \mathbf{H} = \frac{4\pi}{c}\mathbf{J}, \quad \text{div } \mathbf{H} = 0,$$

the displacement current being neglected, we have

$$\nabla^2\mathbf{H} = \frac{4\pi n\epsilon^2}{mc^2}\mathbf{H} = \beta^2\mathbf{H} \qquad \text{.....(24).}$$

It can readily be shown that the magnetic field is confined to a surface layer whose thickness is of the order $1/\beta$, that is about 3×10^{-6} cm. The idea of zero resistance does not come into this theory at all, since the current is determined by the magnetic and not by the electric field. It would therefore be more appropriate to talk of the magnetic moment of a superconducting ring than of the persistent current, and it is, of course, the magnetic field of the ring which is measured. The theory gives an adequate picture of a superconductor in a magnetic field which is less than the threshold field, and is probably on the right lines in focussing attention on the magnetic properties rather than the electrical. It cannot, however, be considered satisfactory when there is an electric field present, as for example in the problem of the gradual restoration of the resistance of a wire by a transverse magnetic field. The great defect in the theory is that it is purely formal, and it is probable that no great advance can be made without a microscopic derivation of equation (23) or an understanding of what happens during the transition to the superconducting state.

REFERENCES

(1) McLennan, J. C. *Reports on Progress in Physics*, **1** (1934).
(2) Silsbee, F. B. *J. Wash. Acad. Sci.* **6**, 597 (1916).
(3) Meissner, W. and Ochsenfeld, R. *Naturwissenschaften*, **21**, 787 (1933).
(4) Casimir, H. and Gorter, C. J. *Physica*, 's Grav., **1**, 306 (1934).
(5) Cf. Abraham, M. *The Classical Theory of Electricity and Magnetism*, pp. 148, 231. (Blackie, 1932.)
(6) Rutgers, A. J. *Physica*, 's Grav., **1**, 1055 (1934).
(7) Keesom, W. H. and Kok, J. A. *Physica*, 's Grav., **1**, 175 (1934).
(8) Kok, J. A. *Physica*, 's Grav., **1**, 1103 (1934).
(9) Sommerfeld, A. *Z. Phys.* **47**, 1 (1928). Cf. Wilson, A. H. *The Theory of Metals*, §§ 1·8 and 5·47. (Cambridge, 1936.)

(10) KEESOM, W. H. and KOK, J. A. *Physica*, 's Grav., **1**, 770 (1934).
(11) HAAS, W. J. DE and CASIMIR-JONKER, M. J. *Physica*, 's Grav., **1**, 291 (1934).
(12) HAAS, W. J. DE, VOOGD, J. and CASIMIR-JONKER, M. J. *Physica*, 's Grav., **1**, 281 (1934).
(13) HAAS, W. J. DE and GUINAU, O. A. *Physica*, 's Grav., **3**, 182 (1936).
SCHOENBERG, D. *Proc. roy. Soc.* A, **155**, 712 (1936).
See also PEIERLS, R. *Proc. roy. Soc.* A, **155**, 615 (1936).
(14) KEESOM, W. H. and KOK, J. A. *Physica*, 's Grav., **1**, 595 (1934).
(15) Cf. MENDELSSOHN, K. Discussion on Superconductivity, *Proc. roy. Soc.* A, **152**, 34 (1935).
(16) HAAS, W. J. DE and BREMMER, H. *Commun. phys. Lab. Univ. Leiden*, Nos. 214 *d* (1931), 220 *bc* (1932).
(17) Cf. WILSON, A. H. *The Theory of Metals*, § 6·61. (Cambridge, 1936.)
(18) BORELIUS, G., KEESOM, W. H., JOHANSSON, C. H. and LINDE, J. O. *Commun. phys. Lab. Univ. Leiden*, No. 217 *c* (1931).
(19) LONDON, F. and LONDON, H. *Proc. roy. Soc.* A, **149**, 71 (1935).

PHOTOELECTRICITY*

(COVERING THE YEARS 1934—1935)

BY G. A. BOUTRY, Directeur, Laboratoire d'Essais du Conservatoire National des Arts et Métiers

§ 1. INTRODUCTION

THE study of photoelectric phenomena presents peculiar difficulties. The first of these is encountered in attempting to define what class of phenomena should be termed photoelectric. Usually, a clear definition and classification of a physical phenomenon can be derived from a thorough examination of bare facts. It is not so in the present case: a clear understanding of photoelectric effects and of their relationship is only to be obtained by starting from an assumption which is not yet accessible to direct experimental proof and which probably never will be.

In a broad sense we may class as photoelectric any direct or indirect consequence of the capture by an electron of the energy carried by one or several photons. This absorbed energy may be used if the electron was initially in a "bound" state, partly to release it and partly to give it kinetic energy; or, if the electron was already in a "free" state, wholly to augment whatever kinetic energy it happened to possess at the instant of capture.

Let us now suppose that this primary process takes place in a solid or in a liquid.

(*a*) Any electron, acquiring sufficient energy to overcome all the internal forces that may oppose its motion, to survive all the collisions it may experience, and finally to cross the potential barrier at the surface, will emerge into the surrounding medium. If the latter is a gas kept at a sufficiently low pressure, the electron may travel far enough to be recognized as such and studied (*photoemission*). The application of an external electric field and the choice of a suitable form of anode will enable the experimenter to capture all the electrons emerging from the illuminated surface. If the photoelectric substance is a good conductor, the permanent current thus generated may be intense enough to be measured, and a *photoelectric cell* is the result. Though many dielectrics and metals have been shown to be photoemissive, in metals alone has the phenomenon been thoroughly studied.

(*b*) Let us now consider electrons which have acquired the requisite energy to pass from a "bound" to a "free" state without being able to cross the surface. On application of an internal electric field, the conductivity of the substance will be found to have increased. This increase will only be perceptible provided the initial number of conduction electrons in the substance was not too great and this explains why, with one possible exception, the phenomenon of *photoconductivity* has only been studied in dielectrics and semi-conductors.

* The warmest thanks of the present writer are due to his friend Mr G. F. Herrenden Harker, who kindly consented to revise the English of the original draft of this report.

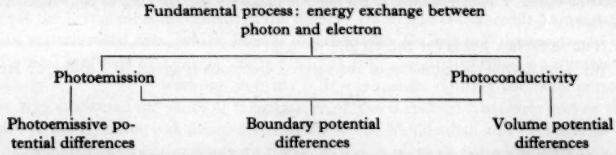
Electromotive forces may be developed as a consequence of these two fundamental phenomena:

(*c*) Electrons are ejected from a photoemissive substance with kinetic energies ranging from zero to a maximum roughly defined by Einstein's relation, and which is of the order of magnitude of a few electron-volts; the consequence is that the contact potential difference between two metallic surfaces maintained in a vacuum and electrically connected is altered by asymmetrical illumination, as can be shown with any commercial photoemissive cell.

(*d*) If a photoconductive substance is illuminated in a non-uniform manner, the electron concentration will not be the same throughout; thus potential differences may arise between two given points or regions of the surface. Provided certain conditions hold (relatively unrestricted motion of the electrons as in a quasi-perfect crystal lattice), these potential differences may maintain a permanent current in a suitably connected external circuit (*volume potential differences* or *Dember effect*).

(*e*) Potential differences may also appear at the boundary layer between two substances whose photoelectric properties are not alike; this has been found to occur when the interface between a metal and a photoconductive substance is illuminated (*photocouples* or *blocking-layer cells*). In a few cases, photovoltaic potential differences (i.e. modification of the e.m.f. of an electrolytic cell by illumination of one of its electrodes) may have the same origin.

The relation between these phenomena is indicated in the following diagram:



Many other phenomena might be included under the term *photoelectric*. For instance, photoionization in gases, which, however, because of close association, is invariably treated as a chapter in the conduction of electricity through gases and also the photoelectric properties of X rays, which, until more is known about the properties of the extreme ultra-violet radiations, are better studied separately. No reference to these is included in the present report.

§ 2. PHOTOEMISSION

(*A*) *Normal photoemission.* Most clean metallic surfaces, investigated in vacuo, exhibit what has been termed *normal photoemission* obeying laws which, for the sake of clarity, are recapitulated here:

(1) The number N of electrons ejected from the surface during any given time is proportional to the number of light quanta n falling on the surface during the same time.

(2) It follows that, if a constant luminous flux W is used, the resulting intensity $i=Ne/t$ is proportional to the flux:

$$i=s.W$$

s may be expressed in amp./lum. or in amp./watt and is called the photoelectric *emission*. If the portion w of the flux absorbed in the metal is alone considered, the modified relation:

$$i=\rho.w$$

holds, in which ρ is termed the photoemissive *yield*.

(3) s and ρ are both functions of the frequency ν of the light used in the experiment:

$$s=\rho(\nu), \quad \rho=\phi(\nu).$$

Starting from zero for a given frequency ν_0 (the *threshold frequency*) s and ρ both increase with ν, no maximum being reached in the visible or near ultra-violet regions of the spectrum.

(4) The velocity of the ejected electrons is independent of W.

(5) The distribution of kinetic energy among the ejected electrons may be defined by a function:

$$\Psi(\nu, V)=\frac{dN}{dV},$$

where dN is the number of electrons ejected with energies ranging from V to $V+dV$ electron-volts. Ψ is *independent of the angle of incidence and the state of polarization of the light*.

The same law holds for the yield.

(6) The spatial distribution of the ejected electrons is given by Lambert's law:

$$\frac{dN}{d\omega}=\frac{N}{2\pi}\cos\phi,$$

where dN is the number of electrons ejected in the elementary cone of solid angle $d\omega$ whose axis is inclined at an angle ϕ to the normal to the surface.

All these laws had been established prior to 1925. Between 1925 and 1932 the effort of workers in this field was mainly concentrated upon obtaining exact and reproducible quantitative data. This proved no easy task, as every physicist who has ever tried to obtain a "clean" surface will appreciate. However, by 1932, s and ν were known with reasonable accuracy for about twenty metals, Ψ had been measured by many investigators with results which were not too divergent. Proposition (6) is the only one for which evidence is still scarce; in spite of noteworthy work by Ives and his collaborators, it still retains a somewhat dubious character[1].

1930 marks a turning point. By that time it was generally conceded that there was no likelihood of developing a plausible theory of photoelectricity on the basis of classical electronics. The new statistics had been so far developed as to provide a starting point for a new theory; and experimental data was available of a quality sufficient to prove or disprove a new physico-mathematical interpretation. There is, admittedly, only one starting point possible: the Fermi-Dirac-Sommerfeld theory

of electronic equilibrium in a metal.* But two paths may be followed; one can either develop a mathematical expression for the spectral distribution functions s_ν and ρ_ν, or one can try to account for the form of the energy distribution curves, in other words compute $\Psi(\nu, V)$. If the calculations are correct, the apparently distinct theories thus evolved will constitute two aspects of a single and complete interpretation of photoemissive phenomena.

Wentzel must be credited with the first serious attempt[2]. His somewhat dubious calculations were revised and elaborated first by Fröhlich[3] and afterwards by Tamm and Schubin[4]. The theory given by the latter cannot be said to be satisfactory; the results are expressed in so complicated a form that it is somewhat difficult to check them against the experimental data. This has been accomplished in one or two instances, but the result was inconclusive as quantitative coincidence was very far from being reached.

Prof. R. H. Fowler was the first to secure this agreement. By restricting his theory to the vicinity of the threshold he greatly simplified his task. It will be remembered[5] that the relation he derived for s_ν is of the form

$$s_\nu = \alpha A T^2 \phi(S),$$

where S is a function of (1) the number of electrons per unit volume, (2) the amplitude of the surface potential barrier, (3) the temperature, (4) the frequency ν. α is a constant probability factor and A is the Richardson constant; in fact, Fowler's expression identifies itself with Richardson's thermionic function if one sets $\nu = 0$; this is the first great achievement of the theory. Its second is the striking agreement with experimental data, which Fowler together with DuBridge[6] demonstrated by graphical methods of analysis and which has since been confirmed by numerous authors[7, 8, 9, 10, 11, 12, 13] working on many different metals. During the same period, the long suspected influence of temperature on the threshold frequency and on the form of the S_ν curve was completely established by Mendenhall and his students[1]. The third achievement of Fowler's theory is that it gives an accurate prediction of that influence. Einstein's equation is valid at no temperature except 0° K.; there is no definite threshold at any other temperature and the S_ν curve tends asymptotically to the axis, the effect growing more and more pronounced with rising temperature. Thus the quantum statistics, born chiefly because Einstein's relation was believed to be absolutely valid, immediately proceed to demolish this rigid character ...a story as old as science itself.

In recent work on the same subject, Suhrmann and Shallamach[13] also obtained results in good agreement with Fowler's calculations. Values for the threshold at 0° K. for about twelve metals have now been computed and agree very closely with the values calculated from the thermionic extraction work functions. Some discussion has arisen about possible secondary effects of temperature, and a new approximation has been put forward by Becker and Brattain[14].

DuBridge and his students have extended Fowler's methods to the computation

* Even this does not provide an entirely satisfactory starting point; it does not take into account the local disturbances caused by atomic fields, disturbances whose effect on photoemission is by no means negligible.

of $\Psi(\nu, V)$, thus completing the theory in the vicinity of the threshold. There again, the agreement between theory and experiment appears to be reasonably satisfactory [15, 16, 17, 18, 19].

Such was the situation at the end of 1934, when DuBridge so admirably summarized the subject [20]. During 1935 a single, but important, attempt was made to solve the complete problem: Mitchell, in three papers [21, 22, 23], gave a general theory, computing both s_ν and $\Psi(\nu, V)$ over a wide range of frequencies and taking into account the influence of temperature. This theory seems very promising and partial confirmation in regard to $\Psi(\nu, V)$ has already been given by Rüdberg [24].

No theory has yet been advanced to take into account the influence of the crystalline state of the surface, which is not insignificant. New evidence in this connexion has been supplied by Underwood, working on single crystals of zinc and copper [25, 26].

(*B*) *Selective photoemission*, most readily studied on alkali or alkaline-earth metals, differs from normal photoemission in that:

(1) The emission depends upon the state of polarization of the incident light, and upon its incidence. It is invariably greatest for light, the plane of polarization of which is parallel to the plane of incidence. The ratio $s\parallel/s\perp$ may be as high as 300 for $i = 60°$. The yield for light $\parallel$ may be much greater than that obtained for normal photoemission.

(2) The s_ν function admits, for unpolarized light or for light polarized $\parallel$, one or more sharp maxima near the threshold.

(3) The spatial distribution of the electrons for light polarized $\parallel$ does not follow the cosine law; its graphical representation, in polar coordinates, appears to be an ellipse having its major axis normal to the illuminated surface.

It has been generally suspected, since 1930, that such a state of things only occurs with surfaces which are not clean *stricto sensu*; the fact that such a phenomenon has only been found with alkali or alkaline-earth metals is explained by the chemical affinity of these elements, and the difficulty involved in thoroughly cleaning their surfaces. (In a few cases where extreme precautions had been taken (1) no selective effect was detected immediately after cleaning, but it slowly appeared in the course of a few days.) The presumption is, then, that selective photoemission can only occur when thin layers of impurities are deposited or absorbed on the metallic surface. It is well known, in fact, that the artificial production of such layers enables cells of great sensitivity exhibiting sharply defined maxima of spectral sensitivity to be manufactured. But, in spite of the extreme industrial importance of such a fact and notwithstanding the work that has been done in many quarters, no progress at all has been made during the period 1931–1935. Nothing is known yet concerning the mechanism by which such an enhanced yield is obtained, though a suspicion prevails that selective absorption of light by these complex layers may be at the root of the matter. Such, at any rate, is Zener's view [27] and such is the starting point of the pro-theory put forward by Lukirsky and Hurgin [28]; though Hlùcká has been developing a theory along somewhat different lines [29, 30]. In the present state of our knowledge, both theoretical and experimental, concerning the structure of surface layers there is little hope of evolving a really coherent and plausible theory

of selective emission. But the spade work is progressing; new, though still fragmentary, evidence has been accumulating[31, 32, 33, 34, 35]: the mathematicians need only wait.

(*C*) *Photoemissive cells.* As regards the sensitivity of photoemissive cells, practically no progress has been made since 1931. The most sensitive cells are still the caesium oxide type developed by Koller and by de Boer and Teves; but reproducibility in manufacture has not so far been realized. Though the cells now sold become progressively more regular, the response of two specimens of the same type under similar conditions may still differ by as much as 50 per cent. This lack of progress is attributable to the difficulties which have just been examined.

Much care has been devoted, of late years, to the study of the photometric behaviour of photoemissive cells. Methods for testing the proportionality of the response of a cell to illumination have been developed in England (at the N.P.L.) by Preston and McDermott[36], in France by Fleury[37] and in Germany by Theissing[38]. The accuracy of the determinations is now better than 1 in 1000 and it has been shown that for commercial cells proportionality rarely holds to such an approximation. The problem is now to manufacture cells of high proportionality and fidelity. In an endeavour to achieve this the present writer has constructed cells with parallel plate metallic electrodes, the anode being only 2 mm. distant from the cathode. The cathode itself is surrounded by a large guard-ring, also coated with a photoemissive deposit. Such a disposition will free the cathode from the disturbing influence of the walls of the cell, and will give a saturation current with an accelerating potential difference of a few volts only, thus minimizing ionization of the residual gas. The testing of these cells is now in progress and encouraging results have been obtained.

Another problem of great interest is the study of the response of a cell to fluctuating light. This has been carried out, for low frequencies and from a photometric point of view, with disappointing results: Talbot's law does not hold for photoemissive cells with more precision than does the proportionality law[39]. Though many papers on the behaviour of gas-filled cells illuminated by rapidly modulated light have appeared[40, 41, 42, 43], no satisfactory theory has yet been formulated which covers all the facts. Campbell and his fellow workers hold that the time-lag does not arise wholly in processes consequent upon a single light impulse, but partly from the interaction of succeeding impulses; should this be confirmed, the improving of gas-filled cells for television and talking-film purposes is likely to prove a very difficult business.

§ 3. PHOTOCONDUCTIVITY

Up to 1920 the record of pure scientific research in the field of photoconductivity roughly amounts to forty years of failure; almost nothing more was known about selenium at the end of that period than was already public property by 1885, a few years after May, in the Valentia telegraph office, first flirted with the fickle element. The explanation of this state of affairs may be that two fundamental propositions had been ignored or insufficiently reckoned with, namely:

(1) that the passage of an electric current through a non-metallic solid medium, however small its intensity and however short its duration, will disturb more or less this medium, whose properties will then keep changing;

(2) that photoconductivity should not be studied as a separate subject, but as a chapter of the more general investigation of conduction through crystals and semi-crystalline materials.

Gudden and Pohl were the first fully to recognize the first of these propositions. Consequently they worked with very weak illuminations of very short duration, choosing for their experiments materials likely to be little disturbed by the conditions imposed, i.e. very perfect single crystals. They soon found that, provided temperature was kept low enough for such crystals to remain perfectly non-conducting in the dark (e.g. diamond at room temperature), the phenomena observed obeyed simple laws, which may be formulated as follows:

(*a*) The quantity of electricity q passing through the crystal is, under any applied potential difference, proportional to the amount of radiant energy w absorbed by the crystal;

(*b*) the charge traversing the crystal is for a given illumination found to increase with the applied field, up to a limit Q (saturation);

(*c*) a *photoconductive yield* may then be defined as the ratio of Q to w:

$$\rho_v = Q/w$$

and may be measured in electron per quantum. If this is done, it is found that exact equivalence prevails, in other words that for a given region of the spectrum one electron is liberated for each quantum absorbed.

All this, and many other important results had been established by 1925. Then a new type of crystal, the prototype of which is rocksalt, came under discussion. Whilst diamond and crystals of the same class will only obey the above laws when they are very pure, pure rocksalt is not photoconductive and will become so only when impurities are present in its lattice (i.e. atoms of sodium in excess). For the currents thus obtained, laws (*a*) and (*b*) are valid, but law (*c*) does not hold.

So far, the properties of about twenty crystalline species under feeble illuminations have been accounted for. But the main body of photoconducting substances (selenium, cuprous oxide, microcrystalline oxides and sulphides, etc.) have failed to manifest any simplicity of behaviour. The large currents one may get from them, the "fatigue" which appears almost at once, the fact that quantum equivalence is often exceeded, all pointed to the existence of a "secondary" effect such as occurs in perfect crystals of the diamond class under intense illumination. But no "primary" current governed by simple laws could be detected.

In 1931, a development in the outlook of research workers became apparent: the second proposition became fully recognized and, at the same time, the principles of the first theory of the electrical behaviour of "semi-conductors" were published by H. A. Wilson. A flood of papers followed, which it would be outside the scope of this report to deal with; mention is only made, in the appended list of references, to those in which the evolution of the theory can be clearly followed[44, 45, 46, 47, 48, 49].

We are only concerned with the impact of this theory on photoconductivity; which may be summarized as follows:

(*a*) In any semi-conductor where the presence of free electrons can be detected (e.g. Hall effect), specific conductivity and photoconductivity are both almost always caused by small amounts of impurities present in the lattice (O in Cu_2O, atoms of metal in excess in alkali or silver halides, etc.): so that no two crystals behave in the same manner.

(*b*) In any such crystal a "primary" current obeying Gudden and Pohl's laws should exist; but

(*c*) this current will be detected only if the "dark" conductivity is sufficiently small, e.g. at suitably low temperatures.

Proposition (*a*) is certainly true in many cases; it is also true that a careful study of single crystals of cuprite at liquid air temperature and under feeble illumination has shown them capable of behaving as rocksalt does at room temperature. As consistent results are still lacking, no quantitative data can be listed here with any confidence, even for cuprite, where the only suspected impurity was oxygen; and all values of the work-function are widely discrepant. Of late years, workers have been trying to account for these variations[50, 51]; the influence of water vapour on the superficial conductivity of cuprite has been proved[52, 53]. Errors concerning its behaviour at high temperature have been rectified[54]; an influence of strong electric fields is now suspected[55]; and the role of another impurity, systematically added copper, investigated[56]. The only conclusion emerging so far is the extreme difficulty of cartesian experiments on such materials.

This seems to have resulted in a general loss of enthusiasm: nobody in 1934 or 1935 has bothered with selenium; one worker records a few observations on mercuric iodide[57]. Happily, Gudden and Pohl's school in Göttingen persevere in the study of the simpler cases (alkali and silver halides); with the aid of the new theory results are slowly accumulating and a few independent workers are taking up the problem[58, 59, 60, 61, 62, 63, 64]. Nothing sufficiently unambiguous to be dealt with in this section has yet appeared.

In a recent series of papers, Majorana[65] claims the existence of photoconductivity of a sort in thin layers of metal. The existence of such an effect seems extremely doubtful, particularly as the author states that it is not instantaneous and the suggestion has been made that the phenomenon is of a purely thermal nature[66]. Little work has been done on photoconductive cells; their characteristics have been investigated in one paper, without disclosing any very arresting results[67]. One more element has been added to the already long list of photoconductive substances, namely boron[68].

§ 4. PHOTOELECTRIC POTENTIAL DIFFERENCES

(*A*) It has already been mentioned that when a photoemissive cell is illuminated, a permanent potential difference is set up between its electrodes; if an ammeter is connected to these, in the absence of any other generator in the circuit, a permanent

current will flow. But, except for cells of very special design (central cathode, spherical anode) this current is almost negligibly small. No recent work has been done on this subject; none seemed necessary.

(*B*) As regards the study of volume potential differences (Dember or crystal photoeffect) the situation is very different. It will be remembered that the effect was discovered by Dember in 1931, using naturally occurring single crystals of cuprite. For some time this remained the only known example but, in 1932, zincblende was shown to exhibit the same phenomenon. Quite recently rocksalt[69], pyrargyrite and stephanite[70] have been added to the list; selenium also, from recent experiments by Holmes[71] may sometimes give a Dember effect.

Early theories of the phenomenon were elaborated, often based on insufficient experimental data. This is now being corrected: cuprite crystals have been frequently studied in 1934 and 1935. It is now known that the thickness of the crystal under experiment is an important factor; strips of different thickness cut from the same crystal will exhibit, if the influence of frequency is investigated, maxima variously located in the spectrum. This is readily explained by the high and selective absorption of cuprite and, when this is taken into account, it is invariably found that the frequency giving the maximum Dember effect coincides approximately with that giving the maximum photoconductivity[72, 73, 74, 50].

This serves to support the view held by most German and Russian workers that the phenomenon is mainly electronic in nature. Recently, as has been done for photoconductivity, a negative and a positive conduction (the former electronic, the latter hole, *not* ionic, conduction) have been considered[50, 76, 77]. A recent modification of the theory by Frenkel[78] is more general in character, and gives an explanation of the photomagnetic phenomenon recently discovered in artificial cuprite. With the possible exception of this work, it cannot be said that any of the many formulae now published give a correct interpretation of the experimental characteristic curves.

A somewhat disquieting fact, now once more confirmed, is that in no case has the crystal photoeffect been produced in artificial crystals of cuprite, however perfect their lattice[79]. This is probably why some workers[80] still hold that the effect is electrolytic and that impurities present in the crystals must be held responsible.

The discovery of the photomagnetic phenomenon was first reported, to the best of the belief of the present writer, in a letter to the Editor of *Nature*, **131**, 725 (1933). Additional data have since been published[81, 82]. Two facts are of interest: (1) The effect has only been observed on artificial microcrystalline cuprite, and (2) the order of magnitude: when a strip of cuprite, illuminated on one of its faces, is placed in a magnetic field perpendicular to the direction of illumination, a potential gradient perpendicular both to light and field is produced. Provided the temperature is that of liquid air, e.m.f.'s of 3 to 5 V. are easily obtained while, with intense fields, 15 V. have been observed[50]. A preliminary theory by Frenkel[83] has been subsequently revised by its author[78]. In view of the startling order of magnitude announced by its discoverers, the lack of interest of the other schools of research

seems somewhat surprising: independent confirmation would be of the highest importance.

(*C*) We now come to the electromotive forces generated in the vicinity of the boundary layer between two photoelectrically dissimilar substances. Cells working upon this principle are called in this report *photocouples* (couples photoélectriques), a denomination which while invoking no hypothesis does, nevertheless, call to mind the important fact that such e.m.f.'s are produced when a difference in illumination is established between two contacts in a complete chain of conductors. A photocouple thus assumes in photoelectricity a role similar to that of a thermocouple in thermoelectricity. This is so far valid that it has been recognized that the presence in the chain of a semi-conductor is not essential[84a]: couples can be made with two different metals separated by a thin film of a non-photoelectric substance. In such a case the cause of the e.m.f. is assuredly electronic; its order of magnitude is very small, as would be expected; the electron concentration is very high in both metals, penetration difficult, speeds and free paths very small. One may hope that some day a coherent theory will give a complete explanation of phenomena so evidently interrelated as Volta potential differences, thermionic emission, and photoelectric boundary e.m.f.'s.

We are still far from this goal. From the starting point of Wilson's theory of semi-conductors and Schottky's hypothesis of the "Sperrsicht", through the work of Fowler, Frenkel, Bornstein and others (1929–1933), and lately in papers by Fowler[85] and Varian[86], the electronic theory of rectifying contacts has often been modified and improved. But it is still unable to face the ordeal of quantitative verification; attempts to secure such verifications have been rare. On one occasion the result was distinctly unfavourable[87]; on another occasion a verification of a rather far-fetched consequence of the theory (the energy necessary to induce an electron to pass from a metal to a semi-conductor across an insulating film should be independent of the nature of the metal) has been so startlingly successful as to raise grave doubts in the present writer's admittedly sceptical mind[88].

It should never be forgotten that the presence of a photoconductive substance in a photoelement implies that, even though the fundamental process may be purely electronic, it may be more or less obscured by secondary currents. Attacking the problem from this angle, Liandrat seems to have been successful in analysing his observations. Working with a Se/Fe couple he proved the existence of a primary current, completely masked after a very short time by a larger and erratic secondary current[89]. His measurements are particularly valuable, since they establish the fact that Einstein's relation holds for the primary process: in this most of Schottky's early views receive confirmation. However, before general assent can be accorded to a result of such importance, confirmatory evidence is desirable and should soon be forthcoming.

The preparation of photocouples has been the subject of a few papers where methods of manufacture are at last well described ([92,93] for Se couples, [94] for Cu_2O couples). Their photometric properties have been studied with increased attention and precision: the relation between current and illumination has occupied

at least six workers during the last two years[95, 96, 97, 98, 99, 100]. The conclusion of these investigations is disappointing; it seems that in no case does the proportionality law really hold, so that precise measurements with photocouples by a direct deviation method should be restricted to a very limited range of illumination, otherwise a correction formula must be employed. But as Elvegaard, for instance, proposes an expression of the form:

$$\phi = i\,[k_1 - k_2 i^{\frac{1}{2}}]^{-2},$$

where k_1 and k_2 are constant for *one* couple used with *one* fixed external resistance, the second solution seems well-nigh hopeless.

Moreover further work makes it clear that no photocouple can compete in respect of constancy and regularity with a photoemissive vacuum cell[101, 102, 103]; the only remaining advantages of the photocouple over its rival are therefore robustness, simplicity of use and a good current sensitivity. On concluding this section, attention may be directed to certain new studies of the dynamical characteristics of photocouples[104, 105, 107] and of the influence of temperature upon them[106]. In the latter connexion it must be borne in mind that the effect is considerable for many $Cu - Cu_2O$ couples and is not negligible even with the selenium types (1 to 3 in 1000 per °C.).

§ 5. APPLICATIONS

Since the advent of photocouples, photoconductive cells have fallen into almost complete disuse; the same is true of photovoltaic elements (electrolytic cells) which have not been considered in this *Report* because nothing definite emerges as yet from the numerous recent publications on the subject. Between the two opponents remaining in the field there should be no real conflict: a summary review of their properties shows conclusively that each type has its own field of application.

In the following summary, the abbreviation *PE* stands for photoemissive cell; *v* is for vacuum; *g* for gas-filled and *PC* means photocouples.

(I) *Electrical characteristics.*

Equivalent resistance R:

PE, *v*: $10^{11} > R > 5.10^8\ \Omega$.

PE, *g*: $10^{11} > R > 2.10^6\ \Omega$ (depending upon voltage applied and illumination used).

PC: dark resistance. Cu_2O: 300 to 1500 Ω according to type. Se-Fe or Se-C: about 10,000 Ω.

Capacity C:

PE: of the order of 10^{-12} fd. *PC*: up to $0{\cdot}1.10^{-6}$ fd./cm².

Self-induction:

PE or *PC*: negligible.

Dark current:

PE, at room temp. in dry atmosphere: from 10^{-13} to 10^{-10} amp. according to type.

PC: 0 when no other generator in circuit.

Output, current:

PE, *v*: from 0·5 to 40 μA./lumen.
PE, *g*: from 0·5 to 120 μA./lumen.
PC: Cu_2O: 10 to 120 μA./lumen. Se-Fe or Se-C: 200 to 400 μA./lumen.

Output, energy:

PE, *v*: $0{\cdot}5.10^{-4}$ to 4.10^{-3} W./lumen.
PE, *g*: $0{\cdot}5.10^{-4}$ to 1.10^{-2} W./lumen.
PC: 10^{-7} to 5.10^{-4} W./lumen.

(II) *Photometric properties.*

Spectral sensitivity:

PE: the positions in the spectrum of the peaks in the curves are roughly as follows:

Cathodes	A.
Na, pure	3400
Na, "sensitized"	7400
K, pure	4300
K, sensitized (glow discharge)	4400
Rb, pure	4700
Cs, pure	6600 (?)
Cs-O film on Ag	7500 3500 2100

PC: Cu_2O couples: position of maximum may be shifted from 5600 to 6300 A.; sensitivity may extend with some, the so-called "back-wall" types, down to 15,000 A. Se couples: around 5900 A.

Relation between current and illumination:

PE, *v*: for a ten-fold increase in illumination, proportionality should hold within 1 per cent; some cells may be much better.
PE, *g*: No general statement is possible (see books listed under 1).
PC: if an approach to proportionality is wanted, ammeters of very low resistance must be used (see § 4).

Fidelity:

PE, *v*: good; current obtained at constant *moderate* illumination may shift about 1 per cent. in a few hours.
PE, *g*: shifts of 15 per cent have been observed with high tensions; if both high tensions and intense illuminations are used shifts of 100 per cent may occur.
PC: "Fatigue" very pronounced under intense illuminations and when measurements last more than a few minutes.

Inertia:

PE, *v*: none.
PE, *g*: none for moderate illuminations and tensions; otherwise very complex behaviour (see section II).
PC: small except when fatigue is conspicuous.

Influence of heat and humidity:

These will not affect *PE*, except to increase the dark current. May modify all the properties of a *PC*.

At least two general remarks applicable to all types of cells must be added: (1) Two measurements can only properly be compared when the light has been distributed each time in the same way over the sensitive area. (2) The same applies to the state of polarization of the light used and to its spectral composition.

It will now be seen that no cell can be deemed a perfect or a universal instrument (and the reader should remember that numerous small defects have not been mentioned here); a research worker who intends to employ a photocell in some new application should never do so without a careful comparison of his requirements with the properties of the various types available.

Generally speaking, a photocouple is a robust instrument, and simple to use; but low resistance ammeters are the only instruments really adapted to them. Amplification is difficult: the energy output is very small and the necessary coupling transformer would require to be very bulky and possess abnormal characteristics. Moreover, because of their high capacity, the impedance of photocouples falls rapidly as the frequency of modulation rises and at 4 kilocycles is about 99 per cent smaller than at 100 cycles. It will give only rough indications if used in connexion with a direct deviation method. Photocouples are chiefly used in industrial photometry (small, portable and cheap illumination-meters for street lighting, photography, matching of lamps in factories, etc.). They are also well adapted for incorporation in cheap and sturdy microphotometers.

On the other hand, a vacuum photoemissive cell should practically always be used either with an electrometer or an amplifier. Though the impedance of a cell-amplifier unit will decrease with increasing frequency, fairly uniform response can be obtained from 20 to 3000 cycles. Provided certain now well-known rules are complied with, very sensitive light-measuring devices can be built[108]. This is so far true that even metrological laboratories have been studying this question: members of the Reichsanstalt staff have recently described two types of amplifiers for photocells[109] and a general article on d.c. amplifiers adapted to photocells has also been published[110]. a.c. amplifiers though sturdier, are less sensitive; one of their great advantages is that the slowly shifting dark-current of the cell is of no importance. An electrometer or d.c. amplifier can detect a current as small as 10^{-17} amp. (corresponding to illuminations of some 10^{-12} lumen.). A 10^{-18} amp. limit has even been claimed on two separate occasions[111]. Needless to say, such units are extremely "temperamental" and may give endless trouble. On the other hand a vacuum cell coupled with an a.c. amplifier and used at constant frequency will detect currents as small as 10^{-16} amp. If one limits oneself to currents exceeding 10^{-13} amp., two stages are ample and, used with a rectifying ammeter, the unit is sturdier than any sensitive galvanometer.

The vacuum photoemissive cell should be used, whenever possible, in connexion with a combined comparison and constant deviation method. Provided this is done, a limit of accuracy of 1 in 10,000 may be reached. Of course, attendant optical troubles, diffused light in monochromators or in objective lenses, the definition of the sources, will generally set another limit, often ten times higher, but that is not the fault of the cell. Thus in homochromatic photometry, a precision of 1 in 1000 is

"easily" obtained; in photographic densitometry, 2 in 1000[113]. In spectrophotometry, errors ranging from 3 per 1000 to 5 per cent have often been stated; but the perfect registering spectrophotometer has yet to be constructed. When it has been, we shall be one step nearer the correct solution of the "problem without a meaning": heterochromatic photometry of discharge tubes. The applications of photoemissive cells and of photocouples to colorimetry have also been numerous. However, the present writer does not propose to examine any recent work in the photometry division; this should be done in another report by a more competent authority.

A problem, the solution of which taxes to the utmost the possibilities of photocells, is that of stellar photometry; the amount of light is extremely minute; the usual atmosphere of the astronomical observatory is cool, but damp so that the apparatus should not be too delicate. For some considerable time, only the Lindemann electrometer has been used in conjunction with photoemissive cells (vacuum or sometimes gas-filled when sensitivity was regarded as more desirable than accuracy). Now, some workers are using d.c. amplifiers[113]. The very sensitive Cs/O/Ag cells have, at room temperature, a considerable dark current and to obviate trouble on this score Hall, at the Yale observatory, has built a photometer in which the cell is kept at $-40°$ C. in a thermos flask[113]. Many other photometers have been described; two recently[114, 115]. About a dozen observatories in Great Britain[114], in France[116], in America[113, 117] and in Germany have recently been equipped with such apparatus. Most of them use a direct-reading method: the light is so scanty that no convenient alternative technique is available.

The photoelectric method of stellar photometry cannot compete with the photographic plate as regards sensitivity; it is not yet possible to study stars smaller than 4 or 5 visual magnitude if measurements accurate to $\pm 0{\cdot}001$ magnitude are aimed at and that is often the necessary order of precision. The photoelectric method has been chiefly directed of late years to the study of variable stars; there it has no rival. Many stars and groups of stars (β Lyrae, ζ Aurigae, W Ursus majoris, and α and β Canis majoris, Procyon, Boss 5070 etc.) have thus been studied at Washburn, Lick, Yale, and Harvard[113, 115, 117, 118, 119, 120]. One difficulty remains concerning the interpretation of results: variable stars fluctuate not only in brightness but also in colour; and it is, in some cases, difficult to know which effect has taken precedence over the other in determining the response of the photocell[121]. It is on this account that a photoelectric method of star colorimetry employing colour filters is now being developed. At Yale alone, about 350 stars over 4 visual magnitude have thus been studied in three approximately monochromatic regions of the spectrum[113]. Smart has, on the other hand, considered the atmospheric extinction factor[122] and given a method for its computation.

Solar and lunar photometry have also necessitated the use of photoemissive cells: Rougier[116] has thus measured the albedo of the moon; using a spherical Cd cell, Bender and Kruger have built[123, 124] a recording photometer for the study of solar ultra-violet radiation. Latarjet and Liandrat have used for the same purpose a Se photocouple combined with an Ag on quartz filter[125].

This brings us to meteorological applications. Here, the quantity of light is generally large and the apparatus should be portable: photocouples have often been tried. Gruner[126] has thus studied the "purple light" of sunset, Byram[127] the transparency of the lower atmosphere. General articles on the use of photocouples in meteorology have been published[128, 129]. It now appears to be generally recognized that, for recording purposes (e.g. of sunshine), the photocouple is not a very reliable instrument: we have seen why.

Before concluding this somewhat lengthy report mention may be made of some of the miscellaneous uses scientists have recently found for photocells. These are astonishingly varied. An extensometer for the measurement of mechanical stress has been built; it weighs 15 grammes and magnifies the deformations 30,000 times[130]. A photo-planimeter has also been designed[131]; opportunities for the employment of photocells have arisen in connexion with temperature measurements[132] and thermoregulation[133]. Numerous papers have been published in which polarimetric and rotary dispersion studies in the ultra-violet spectrum have been made with their aid[134, 135]. Galvanometer deflections may thus be amplified, as is well known, and two new papers have been devoted to this aspect of the subject[136, 137]. A photoelectric method of seismographic recording[138] has even been developed, but there is little point in extending such a list. One can only refer the reader to *Science Abstracts* for a list and an analysis of the score of articles on photon counters and point out the publication of the first extensive paper on the secondary emission auto-amplifying cells[139] which may soon supersede the ordinary thermionic amplifiers in many applications outside the scope of television.

REFERENCES

(1) Recent books on photoelectricity:
GUDDEN. *Lichtelektrische Erscheinungen* (Berlin: Julius Springer, 1928).
HUGHES and DUBRIDGE. *Photoelectric Phenomena* (International Series in Physics: McGraw Hill Book Co., 1932).
BOUTRY. *Les Phénomènes Photoélectriques et leurs applications. Conférences Rapports de Documentation sur la Physique* (Paris: Hermann and Co., 1936).
(2) WENTZEL. *Sommerfeld Festschrift*, p. 79.
(3) FRÖHLICH. *Ann. Phys.*, Lpz., **7**, 103 (1930).
(4) TAMM and SCHUBIN. *Z. Phys.* **68**, 97 (1931).
(5) FOWLER, R. H. *Phys. Rev.* **38**, 45 (1931).
(6) DUBRIDGE. *Phys. Rev.* **39**, 108 (1932).
(7) WARNER. *Phys. Rev.* **38**, 1891 (1931).
(8) GLASCOE. *Phys. Rev.* **38**, 1490 (1931).
(9) DUBRIDGE and ROEHR. *Phys. Rev.* **39**, 99 (1932).
(10) DUBRIDGE and ROEHR. *Phys. Rev.* **42**, 52 (1933).
(11) WELCH. *Phys. Rev.* **40**, 470 (1932).
(12) DUBRIDGE and SMITH. *Phys. Rev.* **46**, 339 (1934).
(13) SUHRMANN and SHALLAMACH. *Z. Phys.* **91**, 775 (1934).
(14) BECKER and BRATTAIN. *Phys. Rev.* **45**, 694 (1934).
(15) DUBRIDGE. *Phys. Rev.* **43**, 727 (1933).
(16) DUBRIDGE and HERGENROTHER. *Phys. Rev.* **44**, 861 (1933).
(17) ROEHR. *Phys. Rev.* **44**, 866 (1933).

(18) DuBridge and Hill. *Phys. Rev.* **46**, 339 (1934).
(19) Brady. *Phys. Rev.* **46**, 768 (1934).
(20) DuBridge. *New Theories of the Photoelectric Effect.* Actualités Scientifiques et Industrielles, No. 268.
(21) Mitchell. *Proc. roy. Soc.* **146**, 442 (1934).
(22) Mitchell. *Proc. Camb. phil. Soc.* **31**, 416 (1935).
(23) Mitchell. *Proc. roy. Soc.* **153**, 513 (1936).
(24) Rüdberg. *Phys. Rev.* **48**, 811 (1935).
(25) Underwood. *Phys. Rev.* **45**, 745 (1934).
(26) Underwood. *Phys. Rev.* **47**, 502 (1935).
(27) Zener. *Phys. Rev.* **47**, 15 (1935).
(28) Lukirsky, P. and Hurgin. *Phys. Z. Sowjet.* **7**, 99 (1935).
(29) Hlučká. *Z. Phys.* **92**, 359 (1934).
(30) Hlučká. *Z. Phys.* **95**, 7, 486 (1935).
(31) Johnson. *Proc. phys. Soc.* **47**, 287 (1935).
(32) Rentschler and Henry. *J. opt. Soc. Amer.* **26**, 30 (1936).
(33) Gei and Truten. *Phys. Z. Sowjet.* **8**, 342 (1935).
(34) Mayer. *Z. tech. Phys.* **16**, 451 (1935).
(35) Gaviola and Strong. *Phys. Rev.* **49**, 441 (1936).
(36) Preston and McDermott. *Proc. phys. Soc.* **46**, 256 (1934).
(37) Fleury. *C.R. Acad. Sci.*, Paris, **199**, 195 (1934).
(38) Theissing. *Phys. Z.* **36**, 683 (1935).
(39) Boutry. *C.R. Acad. Sci.*, Paris, **202**, 1580 (1936).
(40) Leo and Müller. *Phys. Z.* **36**, 113 (1935).
(41) Campbell, Noble and Stoodley. *Proc. phys. Soc.* **48**, 589 (1936).
(42) Roggendorff. *Phys. Z.* **36**, 660 (1935).
(43) Kwartsava and Timofeew. *Tech. Phys. U.S.S.R.* **1**, 469 (1935).
(44) Wilson, H. A. *Proc. roy. Soc.* **133**, 458 (1931); **134**, 277 (1931).
(45) Fowler, R. H. *Proc. roy. Soc.* **140**, 505 (1933); **141**, 56 (1933).
(46) Gudden. *Ergebn. exakt. Naturw.* **13** (Berlin: J. Springer, 1934).
(47) Pohl. *Phys. Z.* **35**, 107 (1934).
(48) Gudden and Schottky. *Z. tech. Phys.* **16**, 323 (1935).
(49) de Boer and van Geel. *Physica*, 's Grav., **2**, 286 (1935).
(50) Joffe. *Les Semi-conducteurs électroniques.* Actualités Scientifiques et Industrielles, No. 202 (1935).
(51) Nasledow and Nemenow. *Phys. Z. Sowjet.* **7**, 513 (1935).
(52) Dubar. *C.R. Acad. Sci.*, Paris, **201**, 883 (1935).
(53) Brauer. *Ann. Phys.*, Lpz., **25**, 609 (1936).
(54) Schottky and Waibel. *Phys. Z.* **36**, 912 (1935).
(55) Borrissow, Kara and Sinelnikow. *Phys. Z. Sowjet.* **8**, 425 (1935).
(56) Waibel. *Z. tech. Phys.* **16**, 366 (1935).
(57) Nix. *Phys. Rev.* **47**, 72 (1935).
(58) Pelz. *S.B. Akad. Wiss. Wien*, **142**, 509 (1933).
(59) Jensch. *Phys. Z. Sowjet.* **5**, 75 (1934).
(60) Grosev. *C.R. Acad. Sci. U.R.S.S.* **4**, 378 (1934).
(61) Kapp. *Ann. Phys.*, Lpz., **22**, 257 (1935).
(62) Tartakowsky and Poddubny. *Z. Phys.* **97**, 765 (1935).
(63) Pohl. *Z. tech. Phys.* **16**, 338 (1935).
(64) Thiele. *Ann. Phys.*, Lpz., **25**, 561 (1936).
(65) Majorana. *Nuovo Cim.* (1933); *R.C. Accad. Lincei*, **18**, 184 and 347 (1933); *Phys. Z.* **35**, 740 (1934).
(66) Etzrodt. *Phys. Z.* **36**, 433 (1935).
(67) Fuchs and Kottas. *Z. tech. Phys.* **17**, 47 (1936).
(68) Freymann and Stieber. *C.R. Acad. Sci.*, Paris, **199**, 1109 (1934).
(69) Grosev. *C.R. Acad. Sci. U.R.S.S.* **1**, 217 (1935).
(70) Barisch. *Ann. Phys.*, Lpz., **21**, 804 (1935).

(71) Holmes. *J. opt. Soc. Amer.* **25**, 326 (1935).
(72) Barbaumow, Shertack and Ziwtschinsky. *Phys. Z. Sowjet.* **5**, 666 (1934).
(73) Joffe, A. and Joffe, A. F. *C.R. Acad. Sci.*, Paris, **199**, 569 (1934).
(74) Kikoin. *Phys. Z. Sowjet.* **6**, 478 (1934).
(75) Barbaumow and Jensch. *Phys. Z. Sowjet.* **9**, 94 (1936).
(76) Mönch. *Z. Phys.* **91**, 264 (1934).
(77) Joffe, A. and Joffe, A. F. *Phys. Z. Sowjet.* **7**, 343 (1935).
(78) Frenkel. *Phys. Z. Sowjet.* **8**, 185 (1935).
(79) Mönch and Stuhler. *Z. Phys.* **91**, 253 (1934).
(80) Deaglio. *Atti Accad. Torino*, **7**, 52 (1934–5).
(81) Kikoin. *C.R. Acad. Sci. U.R.S.S.* **2**, 71 (1934).
(82) Kikoin and Noskow. *Phys. Z. Sowjet.* **5**, 586 (1934).
(83) Frenkel. *Phys. Z. Sowjet.* **5**, 597 (1934).
(84*a*) De Boer and Van Geel. *Physica*, 's Grav., **1**, 531 (1934).
(84*b*) De Boer and Van Geel. *Physica*, 's Grav., **2**, 802 (1935).
(84*c*) Specht. *Z. Phys.* **90**, 145 (1934).
(85) Fowler, R. H. *Proc. Camb. phil. Soc.* **30**, 55 (1934).
(86) Varian. *Phys. Rev.* **46**, 1051 (1934).
(87) Holm and Kirchstein. *Z. tech. Phys.* **16**, 488 (1935).
(88) Kalabuchow. *Z. Phys.* **93**, 702 (1935).
(89) Liandrat. *C.R. Acad. Sci.*, Paris, **198**, 1028 (1934); **199**, 130 (1935); and Thèse, No. 2454, A, 1588 (Paris, 1935).
(90) Roy-Pochon. *C.R. Acad. Sci.*, Paris, **198**, 2083 (1934).
(91) Roulleau. *C.R. Acad. Sci.*, Paris, **202**, 470 (1936).
(92) Barnard. *Proc. phys. Soc.* **47**, 477 (1935).
(93) Görlich. *Z. tech. Phys.* **16**, 268 (1935).
(94) Finck and Fogle. *Trans. Amer. electrochem.* **66**, 271 (1935).
(95) Hamaker and Beezhold. *Physica*, 's Grav., **1**, 119 (1933).
(96) Grundmann and Kassner. *Phys. Z.* **35**, 16 (1934).
(97) Liandrat. *C.R. Acad. Sci.*, Paris, **199**, 1394 (1934).
(98) Elvegaard. *Phys. Z.* **37**, 129 (1936).
(99) Marchal and Marton. *Rev. Opt.* (*théor. instrum.*) **15**, 1 (1936).
(100) König. *Helv. phys. Acta*, **8**, 505 (1935).
(101) Bergmann. *Phys. Z.* **35**, 450 (1934).
(102) Bulian. *Phys. Z.* **36**, 33 (1935).
(103) Frencz and Urbanek. *Rev. Opt.* (*théor. instrum.*) **14**, 317 (1935).
(104) Roe. *Rev. sci. Instrum.* **5**, 441 (1934).
(105) Bloch. *Rev. sci. Instrum.* **6**, 173 (1935).
(106) Roulleau. *C.R. Acad. Sci.*, Paris, **202**, 749 (1936).
(107) Wood. *Rev. sci. Instrum.* **6**, 196 (1935).
(108) Muller and Durichen. *Z. Phys.* **95**, 66 (1935).
(109) Einsporn. *Phys. Z.* **36**, 347 (1935).
(110) Donzelot and Divoux. *J. Phys. Radium*, **5**, 357 (1934).
(111) DuBridge. *Phys. Rev.* **37**, 392 (1931).
Gambetta. *C.R. Acad. Sci.*, Paris, **196**, 342 (1934).
(112) Boutry. *Mesure de densités photographiques par la méthode photoélectrique.* Publications du Ministère de l'Air (Paris, 1934).
(113) Hall. *Astrophys. J.* **79**, 145 (1934).
(114) Smart. *Proc. Roy. astr. Soc.* **94**, 115 (1933).
(115) Calder. *Bull. astr. Obs. Harv.* No. 405 (1935).
(116) Rougier. *C.R. Acad. Sci.*, Paris, **202**, 463 (1936).
(117) Huffer. *Astrophys. J.* **79**, 369 (1934).
(118) Elvey. *Astrophys. J.* **81**, 173 (1935).
(119) Krön. *Astrophys. J.* **82**, 225 (1935).
(120) Fath. *Bull. Lick Obs.* **474**, 115 (1935).
(121) Maggini. *R.C. Accad. Lincei*, **18**, 223 (1933).

(122) SMART. *Proc. Roy. astr. Soc.* **94**, 839 (1934).
(123) BENDER and KRUGER. *Phys. Z.* **35**, 321 (1934).
(124) BENDER. *Phys. Z.* **37**, 107 (1936).
(125) LATARJET and LIANDRAT. *Rev. Opt. (théor. instrum.)*, **14**, 443 (1935).
(126) GRUNER. *Beitr. Geophys.* **46**, 202 (1935).
(127) BYRAM. *J. opt. Soc. Amer.* **25**, 393 (1935).
(128) POLLACK. *Beitr. Geophys.* **41**, 92 (1934).
(129) GRUNDMANN and KASSNER. *Phys. Z.* **35**, 16 (1934).
(130) LEHR and GRANACHER. *ForschArb. IngWes.* **7**, 66 (1936).
(131) GRADSTEIN. *Rev. sci. Instrum.* **5**, 311 (1934).
(132) ZETMANN. *Arch. tech. Messen*, **4**, 116 (1934).
(133) LANGE and VOOS. *Z. tech. Phys.* **15**, 323 (1934).
WALSH and MILAS. *Industr. and Engng Chem.* **7**, 122 (1935).
(134) SHÖNROCK and EINSPORN. *Phys. Z.* **37**, 1 (1936).
(135) BRUHAT and GUINIER. *J. Phys. Radium*, **4**, 691 (1933); *C.R. Acad. Sci.*, Paris, **197**, 1028 (1933).
BRUHAT and GRIVET. *C.R. Acad. Sci.*, Paris, **199**, 718 (1934); *C.R. Acad. Sci.*, Paris, **199**, 852 (1934).
(136) WATTON. *J. sci. Instrum.* **12**, 115 (1935).
(137) DOWNING and HILL. *J. sci. Instrum.* **12**, 91 (1935).
(138) WOLFE. *Rev. sci. Instrum.* **5**, 359 (1934).
(139) ZWORYKIN, MORTON and MALTER. *Proc. Inst. Radio Engrs*, N.Y., **24**, 351 (1936).

OPTICS

Edited by L. C. MARTIN, D.Sc., A.R.C.S., D.I.C.

§ 1. GENERAL INTRODUCTION

By the EDITOR OF THE ARTICLE

THE field of optics is one which has been so industriously cultivated for many years that the majority of the crops which can be grown without undue trouble are fairly well known, at least in the case of ordinary theoretical and instrumental optics; yet fresh workers see old things in new ways, and even the rocky parts, hard indeed to pulverize, continue to attract hopeful people. The progress is thus mainly in matters of detail.

Previous reports in this series have not dealt with ophthalmic optics, but it is a sphere in which real progress is being made, in spite of the lack of adequate optical, physiological, and clinical knowledge on the part of some of those attempting to perform research. The same remark as to progress applies in some measure to subjects such as Colour Vision. Any real advance to be made in future will probably demand a wider preliminary education and experience than has hitherto been thought necessary; and this training must not be shirked. Much has been done already, however, of great practical importance in the study of vision, as will be realized, for example, from the reports beneath.

In several of the sections beneath it has been attempted to survey the progress over several years previous to 1936 in cases where the topics in hand have not been very fully covered in previous reports. This is particularly the case in the sections on Physical Optics and Ophthalmic Optics. The fairly full treatment given to certain topics should make it possible in following reports to give space to other subjects. For example, more attention could be given to details of sundry instruments which receive only passing reference in the following pages, and also to technical problems which escape attention in the usual journals.

§ 2. OPTICAL THEORY

By L. C. MARTIN, D.Sc., A.R.C.S., D.I.C.

A review of this kind is not the place to attempt any detailed discussion of modern advances or attempted advances in the theory of optical instruments. It is not too much to say that the physics of the optical instrument is fairly thoroughly known, and certain outstanding questions have been discussed in recent reports. Theoretical work may be of service in two main ways. It may improve the philosophical approach to the subject so that irrelevant discussions can be completely avoided; secondly, it may facilitate the task of the designer.

The theory of collinear systems certainly offers ground for criticism when used as an introduction to geometrical optics. Von Rohr[1] recently suggested an alter-

native scheme of procedure, following lines initiated by Gullstrand. It resembles in some degree the discussion given by the President of our Society in Appendix B of the *Report on the Teaching of Geometrical Optics*. During 1935 Herzberger[2] has applied his optical tetrality principle to a new development of Gaussian optical formulae. This principle may be briefly stated as follows:

A meridian ray in an axially symmetrical system may be defined by the height $\bar{h}$ in which it intersects a plane perpendicular to the axis, and by its angle σ with the axis in the object space; similarly a ray in the image space is defined by corresponding quantities h' and σ'. Both h and σ are small quantities conforming to the usual paraxial limitations. The quantities $\bar{h}_1$, σ_1; h_1', σ_1' specify an infinitely close neighbouring ray. Then a fundamental invariant connects the object and image spaces, viz.

$$n'(h_1'\sigma' - h'\sigma_1') = n(\bar{h}_1\sigma - \bar{h}\sigma_1).$$

According to the "tetrality" principle, the following substitutions are allowable in all valid equations connecting the above magnitudes:

(1) $n\sigma \to \bar{h}$, $\bar{h} \to -n\sigma$,
(2) $n'\sigma' \to h'$, $h' \to -n'\sigma'$,
(3) $n\sigma \to \bar{h}$, $n'\sigma' \to h'$, $\bar{h} \to -n\sigma$, $h' \to -n'\sigma'$.

(This should not be confused with the "duality" principle published in 1934, which concerns substitutions not subject to paraxial limitations, and which Herzberger uses to discuss the theory of the microscope.) Starting from the above point Herzberger deduces the familiar equations of Gaussian optics.

In spite of all that can be said against the collinear geometry approach to optical theory, it always seems to the writer that the objections are just as unreasonable as those which might be made against the use of scaffolding in the early stages of erecting a building. There are two classes of physicists, one class with a geometrical imagination and another which thinks in symbols and abhors diagrams. It would be hard to say to which class Science owes the more, but they will always demand differing avenues of approach to certain problems.

A note in one of Herzberger's papers refers to the German Standard report DIN 1335 (Deutsche Industrie Normen) which deals with notations in technical optics. This is summarized in the *Zeitschrift für Instrumentenkunde*[3]; it does not attempt to cover the same field as the *Report on the Teaching of Geometrical Optics*, but is almost exclusively concerned with notation. Those familiar with the German optical literature will find little unexpected, but there are a number of useful suggestions, amongst which may be noted the proposed standard use of terms like "magnification" (*vergrösserung*). It is to be restricted to employment in connexion with subjectively used instruments, and will always mean the ratio of the apparent angular size of the image to that of the object. The ratio of conjugate magnitudes in the image and object spaces of optical systems will be called imaging-scale (*Abbildungsmasszstab*). It is a pity that we have not a very good word in English for this purpose.

The note on questions of sign is brief. It recommends that the direction of the light shall be taken as the positive direction of the axis of symmetry, that all formulae shall concern directed lengths, and that conjugate magnitudes shall be measured in the same sense. Every direction not perpendicular to the axis is thus given a direction sense by the choice of the direction of the light.

The report even goes as far as to standardize the symbols and terms to be used for the theory of the primary Seidel aberrations, but this is a dangerous matter unless the whole subject is discussed in detail, and the significance of secondary as well as primary aberration terms is considered.

The actual proposals in this connexion appear to the present writer to be most undesirable.

References

(1) von Rohr. *Z. ophthal. Opt.* **22**, 33 (1934).
(2) Herzberger. *J. opt. Soc. Amer.* **25**, 295 (1935).
(3) DIN 1335. *Z. InstrumKde*, **55**, 184 (1935).

§ 3. OPTICAL INSTRUMENTS (GENERAL)

By L. C. Martin, D.Sc. and B. K. Johnson, D.I.C.

Since the successful casting of the disc for the 200-in. mirror[1] much attention has been directed to the problems concerning big reflecting telescopes, and a number of general reviews[2] of the subject have appeared. One great difficulty experienced, even with a perfectly figured paraboloidal surface, is that the large amount of coma present causes a rapid deterioration of the image away from the axis, so that an experienced observer can detect the deterioration of the image at less than 4 minutes of arc from the axis with the mirror working at about $f/5$. While much can be done by observation of images on the axis alone, problems of measurement often make it desirable to photograph a reasonable extent of the field, and it is then necessary to consider some means of removing the coma. The problem was attacked with some success by Schwarzschild in 1905, and more recently Chrétien[3] published a fresh theoretical solution of the problem of giving such a figure to both a large mirror and a supplementary convex mirror (Cassegranian telescope form) that both spherical aberration and coma (offence against the sine condition) are removed. The equation to the principal section of each mirror appears as a convergent series. Ritchey has figured the mirrors of a 40-in. telescope as closely as possible to accord with Chrétien's equations, but the possibility of the production of satisfactory non-spherical surfaces, even with slight amounts of asphericity, depends very greatly on the possibility of suitable tests, and these are by no means easy to devise.

B. Schmidt used an entirely different method, placing a figured glass plate in the parallel beam before incidence on a spherical mirror, again with some measure of success. So much may be said, by way of introduction, on the situation prior to 1935.

F. E. Ross[4] has now designed a two-lens system to be placed in the convergent

beam near the focus of the parabolic mirror, which removes coma and astigmatism while not introducing excessive spherical or chromatic aberration. The solution is given in formulae enabling the calculation of the radii etc. required in any given case. Systems of this kind have been constructed for the 60-in. and 100-in. mirrors at Mount Wilson.

In estimating the influence of residual amounts of coma on stellar measurements obtained from photographs, an accurate knowledge of the distribution of light in the image when afflicted by coma is desirable. Van Heel[5] has devised a lens producing pure coma for studies of this kind. Reference may be made, however, to papers on similar topics published in this country about 1922–3, in the *Transactions of the Optical Society* and elsewhere.

It was mentioned above that the accurate testing of non-spherical surfaces is difficult. There are several possibilities, illustrated by recent researches. Yvonne Chatelain[6] has described an attempt to check the figure of a paraboloid by using a spherical test plate and observing the interference fringes between the mirror and the plate; an obvious difficulty is that of producing the requisite convex surface to an adequate accuracy. Then C. R. Burch[7] has described further work on the Zernike test and its application to the zonal testing of paraboloids. The advantage in this case is that one obtains with a good "*Z* disc" the equivalent of a perfect spherical convex test plate of variable curvature. Further experiments on these lines by Burch were concerned with the null-testing of paraboloids using the Zernike tests, but the methods were not published in 1935. These methods may be contrasted with the older methods of zonal testing using the Foucault test, and null-testing with the aid of a flat.

Apart from telescope-mirrors, interest is taken in the use of non-spherical surfaces in other directions. Theoretical work by Silberstein[8] facilitates the calculation of the departure of certain surfaces from the spherical locus. A patent[9] in the names of Lee and Kapella mentions the use of conicoidal surfaces for the control of spherical aberration. It seems to be contemplated in certain cases of this kind that the central regions of the figured surface will be given the usual spherical form, and the marginal parts will have the form of a conicoid osculating the sphere in a given zone. Again, non-spherical surfaces are used in a new magnifying lens[10] produced by the firm of Busch.

Purposes of this kind may sometimes necessitate very considerable departure from a sphere, and experience shows that the production of such surfaces free from marked zonal irregularity is not an easy matter. The three points of difficulty are, then, (1) the calculation of the required surface, (2) the devising of a test of adequate accuracy, (3) the technique of production. Of these, it is often forgotten that problem (2) exists, but recent workers have realized it more fully.

Apart from the lens patent of Lee and Kapella mentioned above, there is little to record as regards published work on new designs for photographic lenses, although a number of modifications of existing types are described in patent specifications. The high aperture ratios attained with some modern lenses seem amazing until the relative smallness of the picture often required is remembered; in fact it is

possible to obtain fairly high aperture lenses good enough for cheap amateur cinematograph cameras by almost absurdly simple means as in a system specified by Hansel[11]; this seems no more complex than a Ramsden eyepiece with its lenses moved a little closer together. The appropriate balance between the end and the means is not always observed in optical systems. This last remark suggests a reference to some renewed interest in a topic already discussed to some extent in past years, viz., the correlation between the design of optical instruments and the power or properties of the human eye.

Arnulf[12] has conducted an extensive series of observations on the resolving limit of the eye under varying conditions of illumination and pupil diameter. He applies the results to the specification of the optical system giving the optimum results in any required instance. Flügge[13] has taken up the night-glass problem once again. Although it is doubtful whether any significant change of actual practice is called for by the results of these papers, they do establish some rational basis for the selection of observing instruments, and deserve study by those who have problems in this connexion. Might it be suggested, however, that here again the discussion is conducted on lines similar to those adopted in work published on this subject in this country a good many years ago (1919, and also reviewed in the *Proceedings of the Optical Convention of* 1926).

The field of practical microscopy must claim some notice. Interest continues to be taken in ultra-violet microscopy, and the improved transparency and homogeneity of some of the modern fused quartz led Johnson[14] to design and construct a microscope for use with ultra-violet light of wave length 1990 A. The objective takes the form of a "Mangin" mirror reflecting system, and has been made aplanatic for a numerical aperture of 0·5. Similar designs have also been worked out and constructed for use with $\lambda = 5461$ A.

The modern position seems to be that ordinary ultra-violet microscopy with medium wave-length ($\lambda = 2749$ A.) is not looked upon as presenting any great difficulty in practice. The extension of the range of usable wave-lengths is desirable not only because of the possibilities of increased resolving-power, but because of the varied differential transmission and reflection effects which are capable of much more extended use, for example in the differentiation of biological structures without staining. The great disappointment is the general failure of biologists to realize the possibilities presented by the new optical methods of research, and their apparent unwillingness, with a few distinguished exceptions, to acquire the necessary training and technique.

The introduction, a year or two ago, of the "Homal" projector lenses for photomicrography had the object of flattening the image field. With a similar end in view Köhler[15] has described a series of objectives for microphotography and micro-projection to be used without projector lenses. Systems of this kind usually take the general design of small photographic lenses, and constructions similar to the Cooke lens and the Tessar are recognizable. A patent[16] under the name of Zeiss describes a microscope objective in which the front components, of more or less orthodox form, are given a positive Petzval sum, while the back components,

comprising a divergent member between two convergent members, have a negative Petzval sum. This construction also aims at achieving a flat field.

So far as theoretical attention has been given to the subject of microscopic vision, the tendency seems to be to explore the possibility of using the instrument more as a kind of interferometer and less as a picture-making device. A contribution to this end is the application of the Zernike illumination to the microscope[17]. For this purpose the illuminating beam may be given an extremely small aperture; then, instead of blocking the direct light as in dark-ground illumination, it is passed through a phase-retarding layer of refracting material. This brings out a differentiation of colour and brightness (due to interference) in the "image" according to the optical thickness of the object. While this procedure entails much less precision than in the application of the phase-retarding disc method of mirror or lens testing, it does seem to offer a fresh method of some value for the differentiation of organic structure without staining.

There are a good many other topics which might be discussed under this heading; strenuous efforts continue to be made in photo-grammetric methods of surveying; kinematography and television systems continue to occupy a big share of attention amongst certain groups of workers, but the tendency is against publication of results. It may be suggested, however, for future issues of this report, that specialists might be invited to contribute explanatory articles on such matters, since progress in the ordinary branches of optics is not so rapid that they must necessarily be treated in every issue.

References

(1) McCauley. *J. Soc. Glass Tech.* **19**, 156 (1935).
(2) Jones. *J. sci. Instrum.* **12**, 37 (1935).
(3) Chrétien. *Rev. Opt. (théor. instrum.)*, **1**, 13 (1922).
(4) Ross. *Astrophys. J.* **81**, 156 (1935).
(5) van Heel. *Physica, 's* Grav., **2**, 63 (1935).
(6) Chatelain. *Rev. Opt. (théor. instrum.)*, **14**, 41 (1935).
(7) Burch. *Mon. Not. R. astr. Soc.* **95**, 548 (1935).
(8) Silberstein. *J. opt. Soc. Amer.* **25**, 340 (1935).
(9) Lee and Kapella. British Patent 435,149 (1935).
(10) Flügge. *Z. InstrumKde*, **55**, 193 (1935).
(11) Hansel. British Patent 431,726 (1935).
(12) Arnulf. *C. R. Acad. Sci.*, Paris, **200**, 52 (1935).
(13) Flügge. *Z. InstrumKde*, **55**, 367 (1935).
(14) Johnson. *J. sci Instrum.* **11**, 384 (1934).
(15) Köhler. *Z. InstrumKde*, **55**, 407 (1935).
(16) Zeiss. British Patent 438,860 (1935).
(17) Zernike. *Z. tech. Phys.* **16**, 454 (1935).

§ 4. PHYSICAL OPTICS

By F. Simeon, D.Sc.

Absorption. In view of its importance in the investigation of molecular structure, the subject of absorption is receiving increasing attention, a point of particular interest being the structure, if any, of the absorption band, which is discernible

when the substance is in a suitable state, i.e. vaporized or at low temperature. To observe such structure, a continuous light source is necessary. Such a source for the visible region of the spectrum is easy of attainment. For the ultra-violet, however, until recently this has not been the case. It is true that in the hands of a few experimenters the aluminium spark under water has yielded good results, but the electrical equipment is costly and complicated, and there are other disadvantages.

Under suitable conditions, many gases—in particular hydrogen—yield a continuous spectrum in the ultra-violet, while the simplicity and cleanliness of a gas discharge tube render it attractive. Such hydrogen lamps have been described by Bay and Steiner[1], Chalonge and Lambrey[2], Kistiakowsky[3] and several others[4,5,6,7], while recently one has been described by Munch[8] which can be excited at low voltage and requires only 5 per cent of the power needed by tubes excited by high voltage and in addition needs no water cooling.

It is recorded by Kistiakowsky that the spectrum extends beyond the limit of transmission of fused silica, and it is thought to extend into the vacuum region to 1600 A. or even further. As practically every substance shows strong absorption in this region, it is likely to become of the greatest interest—particularly for those substances which are transparent throughout the visible and ordinary ultra-violet regions.

In absorption, as in most fields of spectroscopy, the microphotometer plays a conspicuous part. Many instruments have been described which, although based upon the same principles of working, differ from one another in many points of detail. One of the best known, due to Moll[9], employs a thermopile as detecting element. While subsequent designs have incorporated some form of photo-cell, and the more recent ones a cell of the copper-oxide or rectifier type, one has lately been described by Leighton, Smith and Henson[10] incorporating a radiometer. Another feature, which has tended to become classic in design since the introduction of the Moll instrument, is movement of the plate under investigation by means of a micrometer screw, and of the recording plate or paper by some mechanism coupled to the screw.

Here again there have been departures from this arrangement in instruments described in the last two years. Weigle[11] used a galvanometer in which the axis of the suspension was horizontal while the whole of the galvanometer was rotated about a vertical axis, the recording plate being fixed. Woodward and Horner[12] have described an instrument in which both the plate under examination and the recording plate are fixed while the whole photometer unit comprising exploring slit, photocell, electrometer and optical train is moved, the motion being gravity-operated against an oil leak. For the purpose of investigating small changes of contrast Sears[13] superimposes a vibration of small amplitude upon the translation of the plates, and an a.c. amplifier is connected to the cell.

A simple non-recording instrument has been described by Follett[14] for determining the match points in divided beam spectrophotometry. For many years the Hilger sector photometer has been the standard method of measuring absorptions in the ultra-violet, the principle being that the spectrum taken through the absorbing substance is compared with a contiguous spectrum cut down in intensity in a

known proportion. At the match points of these spectra the absorption of the substance is therefor known. Other photometers[15, 16], of which the Spekker is the latest, working on the same principle, have been introduced from time to time. A similar arrangement[17] recently described by Harrison and McDonald is the notched echelon cell. It is for use in conjunction with such apparatus that the instrument described by Follett is designed and it can also be used for investigation of emission spectra by using one only of the two photocells.

A novel form of microphotometer has been described by Harrison[18]. It consists of a device which can be attached to a moving plate comparator which automatically records the passage of the density maximum of a line, computes the wavelength and records this value photographically. A gain of from 10-fold to 200-fold in speed of measurement is claimed. Another instrument worthy of mention is a recording photoelectric spectrophotometer by Hardy[19], improvements in which have lately been described. The sample and a standard magnesium carbonate surface are alternately presented to a photocell. Any difference of intensity produces an alternating current which can be amplified and made to operate a motor which adjusts a sector until balance is obtained. A complete absorption curve is traced in about 30 sec.

Diffraction. The ruled diffraction grating being still in some respects unrivalled for spectroscopy it is not surprising that papers continue to appear dealing with their production and theory. The formation of sharp images of X-ray lines by narrow plane gratings used at nearly grazing incidence has been dealt with by Porter[20] and more recently by Bisacre[21], while Mack, Stehn and Edlen[22] have considered the dimensions to give maximum resolution. Siegbahn and Magnusson[23] described a machine built at Upsala for the production of such gratings ruled on glass, while more recently[24] concave glass gratings have been ruled in the same laboratory with which valuable research is being carried out in the extreme ultra-violet region. Stone[25], who has had great experience with Grayson's ruling engine at Melbourne, has suggested a number of improvements which might be introduced into ruling engines.

While glass is now a favourite material for ruling gratings for the investigation of very short waves, speculum metal gratings are most usual for use in the visible and near ultra-violet. The reflecting power of speculum, however, is not so high as could be desired, particularly in the ultra-violet. The process of evaporation of metals in vacuum has been worked out by Ritschl[26], Strong[27] and others, and a film of aluminium so prepared provides a way of increasing this reflection, Focke[28] having found increases of 45 per cent, 80 per cent and 170 per cent at 4077 A., 3341 A. and 2536 A. respectively.

Polarization. In common with most fields of optics, increased accuracy has been sought in polarimetry by the introduction of photocells. Bruhat[29] has described arrangements in which such devices have been introduced. While the accuracy claimed does not exceed that of good visual determinations and there may be uncertainty as to the mean wavelength employed in a determination, it may well be that such methods will lead to elimination of fatigue, more particularly in making measurements at the blue end of the spectrum. Circular dichroism is receiving

increasing attention by chemists. The arrangement described by Kuhn[30] seems to provide a practical set up for work in the ultra-violet. It includes a split Rochon prism as polarizer and a Fresnel rhomb of fused silica. Some form of compensator is substituted for the last item by other experimenters, a recent example being that of Mathieu[31].

Another interesting example of the use of polarized light is the determination of refractive index from a measurement of Brewster's angle. This enables indices of crystals above 1·8 to be measured (Maier[32]), but also provides a convenient way of dealing with materials such as paint which must be examined as thin layers and are yet opaque.

Refraction. Glass still remains of paramount importance to optical instruments, so that the causes of variation of refractive index throughout a block of glass are of great interest. That considerable changes could be occasioned by rapid cooling was shown some years ago by Twyman[33] and Simeon, who investigated also the recovery made by annealing at various temperatures. The lowering by rapid cooling has recently been investigated by Klein[34] who has observed a partial recovery after the lapse of a number of months, and predicts that 50 years would be required to effect complete recovery.

A similar effect in plate glass appears to have been independently discovered by Hopfield and Watkins[35].

References

(1) Bay, L. and Steiner, W. *Z. Phys.* **45**, 56, 337 (1927), and **59**, 1–2, 48 (1929).
(2) Chalonge, D. and Lambrey, M. *Rev. Opt. (théor. instrum.)*, **8**, 332 (1929).
(3) Kistiakowsky, G. B. *Rev. sci. Instrum.* **2**, 549 (1931).
(4) Darbyshire, O. *J. sci. Instrum.* **10**, 322 (1933).
(5) Braddick, H. J. J. *Proc. Camb. phil. Soc.* **30**, 355 (1934).
(6) Stevens, D. S. *Rev. sci. Instrum.* **6**, 40 (1935).
(7) Watson, W. H. and Hurst, D. G. *Canad. J. Res.* **13**, Sect. A, 19 (1935).
(8) Munch, R. H. *J. Amer. chem. Soc.* **57**, 1863 (1935).
(9) Moll, W. J. H. *Proc. phys. Soc.* **33**, 207 (1921).
(10) Leighton, P. A., Smith, S. and Henson, F. C. *Rev. sci. Instrum.* **5**, 431 (1934).
(11) Weigle, J. *Rev. sci. Instrum.* **4**, 595 (1933).
(12) Woodward, L. A. and Horner, R. G. *J. sci. Instrum.* **12**, 17 (1935).
(13) Sears, F. W. *J. opt. Soc. Amer.* **25**, 162 (1935).
(14) Follett, D. H. *Proc. phys. Soc.* **47**, 125 (1935).
(15) Lewis, S. J. *Trans. chem. Soc.* **115**, 312 (1919).
(16) Twyman, F. *Trans. opt. Soc.*, Lond., **33**, 9 (1932).
(17) Twyman, F. *Proc. phys. Soc.* **45**, 1 (1933).
Harrison, G. R. *J. opt. Soc. Amer.* **24**, 59 (1934).
McDonald, E. *J. Franklin Inst.* **221**, 103 (1936).
(18) Harrison, G. R. *J. opt. Soc. Amer.* **25**, 169 (1935).
(19) Hardy, A. C. *J. opt. Soc. Amer. and Rev. sci. Instrum.* **18**, 96 (1929); *J. opt. Soc. Amer.* **25**, 305 (1935).
(20) Porter, A. W. *Phil. Mag.* **5**, 1067 (1928).
(21) Bisacre, F. F. P. *Proc. phys. Soc.* **47**, 948 (1935).
(22) Mack, J. E., Stehn, S. R. and Edlen, B. *J. opt. Soc. Amer.* **22**, 245 (1932).
(23) Siegbahn, M. and Magnusson, T. *Z. Phys.* **62**, 7–8, 435 (1930).
(24) Siegbahn, M. and Magnusson, T. *Z. Phys.* **95**, 3–4, 133 (1935).
(25) Stone, W. *J. sci. Instrum.* **11**, 241 (1934).

(26) RITSCHL, R. *Z. Phys.* **69**, 578 (1931).
(27) STRONG, J. and C. H. CARTWRIGHT. *Rev. sci. Instrum.* **2**, 189 (1931).
(28) FOCKE, A. B. *Phys. Rev.* **49**, 650 (1936).
(29) BRUHAT, G. and GUINIER, A. *Rev. Opt.* (*théor. instrum.*), **12**, 396 (1933).
(30) KUHN, W. and BRAUN, E. *Z. phys. Chem.* B, **8**, 445 (1930).
(31) MATHIEU, J. P. *Ann. Phys.*, Paris, **3**, 371 (1935).
(32) MAIER, W. *Z. InstrumKde*, **55**, 49 (1935).
(33) TWYMAN, F. and SIMEON, F. *J. Soc. Glass. Techn.* **7**, 199 (1923).
(34) KLEIN, NUITA. *C.R. Acad. Sci.*, Paris, **200**, 1848 (1935), and **201**, 339 (1935).
(35) HOPFIELD, J. J. and WATKINS, G. B. *J. opt. Soc. Amer.* **26**, 24 (1936).

§ 5. OPHTHALMIC OPTICS

By WILLIAM SWAINE, B.Sc., F.Inst.P.

Ophthalmic optics is concerned with the study of aided and unaided vision together with the methods and instruments used in measuring and correcting visual deficiences—principally of acuity, refraction and binocular vision. Despite the wide individual interest of the subject matter and the important social and physiological consequences of visual errors, its systematic academic study has been comparatively neglected. From various points of view the co-operation of the physicist is increasingly in demand; this is partly confirmed by the recent growth of classes in the subject at technical colleges.

The usual school text-book on light scarcely places the matter in correct perspective, and is even misleading when describing defects of vision. The Physical Society's report[1] (1934) on the teaching of geometrical optics helps to bridge the first gap between scholastic and technical conceptions.

The present report covers the period from 1928 to the present time, connecting up with the Optical Convention (1926) and the separate existence of the Optical Society (1932), and deals with ophthalmic optics as defined above. Defective vision is only briefly considered in the admirable reviews provided by the *Joint Discussion on Vision*[2] (1932) and the *Report of Progress in Physics* (1934).

Ophthalmic optics. Teaching and general theory. The reduced form of the conjugate focus relation[1,3]

$$L+F=L',$$

where

$$L=n/l;\ \ F=n'/f';\ \ L'=n'/l',$$

is most useful in ophthalmic problems, especially when applied as a "step along" addition of reduced vergence, starting from the object. It includes the single surface refraction when $F=(n'-n)/r$, and reflection when $n'=-n$. Capital letters describe vergence, whilst the convenient dioptre unit is the reciprocal of a reduced distance measured in metres. Convergence is numerically positive.

By continuing the "step along" process the transfer of origin for a second action is effected and written down in sequence thus:

$$(L+F)/\{1-\delta\,(L+F)\}=L_2 \quad \text{for } \delta \text{ small}$$

or

$$(L+F)\,D/\{D-(L+F)\}=L_2 \quad \text{for } \delta \text{ large,}$$

where

$$\delta=d/n'=1/D.$$

Such transfer is commonly required for thick lenses (vertex relations), the separation of a lens from an eye, and the still greater separation occurring in cases such as in retinoscopy.

Besides the focusing properties of elements the prismatic deviations of rays are commonly required. The conjugate focus relation expresses a special case of deviation ($hL+hF=hL'$) where hF expresses the prismatic deviation at an incident height of h, in prism dioptres when h is in centimetres and F in dioptres. The lateral magnification relation,

$$yL=y'L', \quad [\text{cf. } n\alpha]$$

is usefully expressed in a step-along order of lateral displacement thus:

$$yL/(L+F)=y'.$$

The theory of neutralizing movements, decentred lenses separated from the eye, Holmes' stereoscope, etc. can immediately be written in one line by such step-along methods. Imagine the case of an eye looking normally, through a lens of power F, decentred c, at an object (placed at L and decentred from the eye axis by z with eye-lens separation D as defined above); starting from the object, the apparent angular displacement of the object at the eye is

$$[\{(z+c)\,L/(L+F)\}-c]\,(L+F)\,D/\{D-(L+F)\},$$

which simplifies to $$(zL-cF)\,D/\{D-(L+F)\}.$$

By a similar process almost as direct, it can be shown that the apparent angular movement ϕ of the reflex in retinoscopy is given by

$$\phi=2\theta D^2/(D+R)\,(L+F-D),$$

where L, F and D have the same significance as before, F is the catoptric power of the retinoscope mirror, R the refraction of the observed eye, and θ the rotation of mirror producing the apparent reflex movement.

Ophthalmic lenses: specification. The static refraction of the eye is the dioptral expression of the far point distance, and is generally expressed as the power of the correcting lens which forms an image of the distant object at the far point. Myopia and hypermetropia are measured in the same way. As lenses are placed so that the back vertex of the lens just clears the eyelashes, the back-vertex designation of lenses has gained ground, a numeration which accidentally agrees closely with the nominal marking of symmetrical trial lenses (Optical Society 1908 Standards)[4] so extensively used. The suggested plano-form with reduced diameter and thickness, marked according to principal power (Ophthalmic Standards 1927)[5], has not achieved popularity, probably because of the unsymmetrical form and because it does not give vertex measure directly. Experience with the prescription of meniscus lenses has justified the precautions taken to ensure back vertex equality in distance vision. Optometers and refractometers for objective examination of ocular refraction are now calibrated for vertex power, and have been found to give reliable results. Some forms of lens battery with widely separated elements cannot give back vertex power by simple addition, although intended to do so, and users have

had to resort to a final trial with lenses, or have measured emergent vergence with vertex-calibrated focimeters. These focimeters, incidentally, are rapidly supplanting neutralization methods.

Whilst vertex designation is valuable, it is not fully realized by professional men that complete equivalence is not ensured[6,7]. A test lens should reproduce, as far as is practically convenient, the focal and magnification properties of the subsequent correction, and this can only be effected by equality of principal power and coincidence of principal planes. The greater attention now paid to dynamic refraction and to the sizes of ocular images demands the further consideration of some agreed procedure, among a number of possible simple solutions, to ensure complete equivalence.

Ophthalmic lenses: theory and design. The theory of decentred spherical and astigmatic lenses has been extended to include separations and the result applied to binocular vision[3,8,9]. An immediate practical result has been the introduction of a technique in the prescribing and design of so-called prism-controlled bifocals[3].

Best-form design[1c,3] of meniscus lenses still aims at minimizing astigmatism of oblique pencils or at placing the circles of least confusion on the far point sphere. The compromise solution in the case of astigmatic lenses commonly aims at equalizing the oblique errors in the principal meridians. Whitwell has extended his calculations to fused lenticular cataract lenses, bifocal lenses and lenses for near work application[10]. The old prejudice against toric lenses is dying, largely because back vertex numbering renders prescription safer. There still remains a number of isolated cases where intolerance of best form correction is shown.

Innovations of lens shape, the elevation of the spectacle side above the horizontal plane of the eyes and the Rheinberg hinged spectacle may be noted. Unlike its historic predecessor, the hinged lens is concave, is in action for distance vision and is flicked out of action for near work. Contact lenses, through the efforts of Heine[1c,11] and Messrs Zeiss, have made more rapid progress, being produced to correct refraction, remove irregular corneal refraction and fit on the sclerotic coat. Telescopic spectacles also are being extended in application[1c,12]. It is too early to assess the clinical value of lenses for equalizing the sizes of ocular images, although their adaptation is growing.

Static refraction. Generally speaking, static refraction method has undergone little alteration; it still occupies the primary position in routine. The number of designs of objective refractometer has increased[36], with little alteration of principle. Streak and focal line retinoscopy have been re-introduced with doubtful advantage, the general aim being to restrict observation to one meridian at a time[19]. Lindner has developed a special system for studying the effects of cylindrical lenses. Of the subjective developments the chromatic tests of Brown (duochrome-bisecting the chromatic interval) and Roessler (cobalt filter) deserve special mention[35]. The Maddox V test figures commonly on test charts, and the Raubitschek figure is achieving popularity[33]. A flux of tests using the Scheiner principle under various names, and probably best described as subjective-shadow methods with stationary

or moving observation, has culminated in the velono-streak test of Lindner; as clinical tests these are somewhat complicated, and in many cases are used only to confirm other tests[35].

These different methods of estimating refractive errors give significant differences (standard deviation $\pm 0{\cdot}5$) which are partly explained in terms of spherical aberration[13]. Objective refractometer readings in general give the nearest approximation to a mean; ordinary subjective methods tend to give a relatively higher hypermetropia measure; retinoscopy is more variable according to the examiner. Bentall and Diprose[14] endeavoured to locate the position of the retinal-choroidal surface which is effective in retinoscopy, and compared the discrepancy between subjective and retinoscopic findings with chromatic and spherical aberrations.

They attribute the major difference to spherical aberration. Various evidences suggest that the simple geometrical theory of refractive errors has reached a limit of utility, especially of the lower errors.

Using retinoscopy alone, Stine investigated the apparent refraction in five regions of the pupil, classifying eyes into positive, negative, and mixed aberration groups[15]. The zonal differences of refraction (dilated pupils) amounted to as much as $7{\cdot}5$ *D*, positive aberration being the commonest. Whilst Ferree and Rand's examinations of eccentric refractions[16] with the Zeiss refractionometer are not immediately comparable with the above central refraction observations, some of their findings for small obliquities may be noted (one dioptre variation within 10° of the central line). This work was primarily concerned with problems of peripheral perception, and demonstrated the variation of relative positions of the focal lines for oblique refractions up to 60° in the horizontal plane. Vertical meridian refractions tended more uniformly to hypermetropia; horizontal meridian refractions varied between relative increase of myopia, decrease of myopia, and asymmetry. It was noted incidentally that prolonged fixation with or without cycloplegia produced increased myopia, especially during peripheral fixation.

Dynamic refraction. Ames and Gliddon[17] have made a systematic study of various ocular functions using a half-silvered open reflecting stereoscope. They established that the eyes do not in general focus accurately for a plane of print: this destroys the common interpretation of many dynamic tests, including Donders's relative accommodation.

When images which are in part similar and in part dissimilar are formed on the two retinae, similar parts may fuse but the dissimilar parts are in general not fused, the disparity being influenced by accommodation and versions of the visual axes. The phenomenon has been described by Ames and Gliddon as retinal slip, the crossed type of diplopia of the dissimilar parts—exophoric slip—can be made to vanish by increased accommodation or by relative divergence of the axes. A high degree of correlation exists between the phoria of disassociated eyes and the version required to correct slip. Relative accommodation and convergence is considered as non-existent: for each vergence condition there is a particular condition of accommodation and retinal slip. Fincham also records refractive variation with altered vergence, the object distance being maintained constant. Ames, Gliddon and Ogle

have extended the study of slip, phoria, retinal asymmetry, and stereoscopy[18]; the clinical application to the diagnosis and the treatment of aniseikonia (unequal size or shape of ocular images—distinct from retinal images) has resulted in the alleviation of symptoms in which ordinary refraction methods have failed.

Dynamic retinoscopy[19] is still largely empirical, the existing theories being inadequate. The same can be said of dynamic refraction generally. Extraneous factors may be noted other than the inadequacy of the simple geometrical theory.

Visual acuity. The use of letters for the measurement of visual acuity has been re-affirmed by various experimenters. A paper by Coates, chiefly valuable for its experimental results, deals with the careful measurement of the relative acuities of a variety of letter and geometrical designs[20]. Acuities for letters were found by two methods, (*a*) simple recognition and (*b*) recognition and drawing component parts when placed at different angles. The latter was about seventy per cent of the former, varying with design. It was suggested that letters chosen for acuity measurement should show the least difference between the two tests, together with equal legibility. There is a remarkable rise of apparent acuity when two letters of equal legibility are exposed for simultaneous comparison. The range of variation of acuity values for different shapes tested was from 4·0 to 0·5, roughly in order of complexity of design, and can be reasonably interpreted as recognition of brightness difference in the retinal image, the mental factor exerting a smaller influence than generally supposed.

Luckiesh and Moss have gathered together the results of extensive experiments on acuity, exposure time, brightness and contrast in their book, stressing the joint partnership of illumination and vision[21]. Most experiments on the illumination-acuity relationship use the Landolt *C* or grating acuity which are now fairly well correlated with Snellen acuity for normal vision, but the results of Hamblin and Winser on the grating acuity in different meridians may be noted with the indication that more attention must be paid to a physical theory of the image[22]. Closely related are the general questions of the mechanism of vision[23], effects of stimulus distance[24], peripheral vision[25] etc., (with immediate physical interest) as well as the more socially important questions connected with occupation, eyes of school-children, motorists' vision etc.[26].

From the clinical point of view the greatest progress has probably been made in the diagnosis, and treatment by optical methods, of amblyopia and squint. The chief physical interest is to be found in the designs of different instruments. Some of these have been elaborate and in many cases introduced without adequate investigation of the physiological and psychological groundwork. Later tendencies are towards simplification. It is difficult to obtain classified statistics describing the results of different diagnoses and treatments, although the work of Cantonnet and Filliozat[27] is gaining many adherents. Instruments fall into three categories; diaphragm separation of the eyes, the reflecting stereoscope (e.g. amblyoscope developed to the mechanically more perfect synoptophore), prismatic stereoscope with prisms or spheroprisms, as in the Holmes' adaptation of the Brewster instru-

ment[28]. A most recent development is an attempt to introduce clinical instruments for studying the movements of the eyes during reading[29].

Space does not permit of a review of work in the physiology of binocular and stereoscopic vision. In addition to matters discussed in the Physical Society's *Joint Discussion on Vision*[30], one may recall the work of Ames, Ogle and Gliddon[18] on phoria and the horopter and Adamson on the cyclo-rotational movements of the eyes[32].

With regard to the mechanism of accommodation Fincham[31] has demonstrated the form of the crystalline lens in the living eye as though a horizontal section had been made, using a slit lamp and ingenious photographic re-projection.

Ophthalmic instruments. Reference has been made in the previous sections to instrumental aids for refraction, acuity and binocular vision. Fincham's coincidence optometer deserves special mention[36]. Steady progress has been maintained in the improvement of the purely diagnostic instruments such as the perimeter, scotometer, slit lamp, opthalmoscope, ophthalmometer, etc. Reference to the informative catalogues, particularly of Messrs Zeiss and Messrs Bausch and Lomb, reveal the extent to which ophthalmic examinations are widely becoming mechanized. The weakest element in this process of mechanization is to be found in those instruments for refraction which are placed immediately before the eyes of an examinee. They do not in general give a measure under natural pupil conditions, whilst some are optically inaccurate.

References

Teaching

(1) *Report on Teaching of Geometrical Optics.* The Physical Society (1934). In addition to the classified text-books enumerated the following may be added: (*a*) Fincham, W. H. A., *Optics* (1934). Group 1, Case (1). (*b*) Duke-Elder, W. Stewart, *Text-Book of Ophthalmology* (1932). Group 1, Case (2). (*c*) Rohr, M. von and Boegehold, H., *Das Brillenglas als Optisches Instrument* (1934). Group 1, Case (1).

(2) *Report of a Joint Discussion on Vision.* Physical Society (1932).

(3) Emsley, H. H. and Swaine, W. *Ophthalmic Lenses* (1935).
Martin, L. C. *Applied Optics* (1930).

Ophthalmic lenses

(4) "Report of the Optical Society Standards Committee on the standardization of trial cases." *Trans. opt. Soc.*, Lond., **9**, 138 (1907–8).

(5) "Standards of ophthalmic products." *Refractionist*, **16** (Supplement), Nov. 16, 231 (1927).

(6) Swaine, W. "A suggested standard trial case and simplification in ophthalmic policy." *Trans. opt. Soc.*, Lond., **25**, 2, 73 (1924).

(7) Smith, T. "Note on use of lenses in series for sight testing." *Trans. opt. Soc.*, Lond., **29**, 5, 243 (1928).
Smith, T. "Back vertex power of a combination of lenses." *Trans. opt. Soc.*, Lond., **26**, 1, 31 (1924).
Smith, T. "Theory of neutralization." *Trans. opt. Soc.*, Lond., **26**, 1, 38 (1924).

(8) Tomkins, J. A. "On the prismatic effect of lenses in near work." *Optician*, **80**, 2080, 427 (1931).

(9) Whitwell, A. "The Stokes lens." *Optician*, **76**, 1961, 153 (26 Oct. 1928).

WHITWELL, A. "Obliquely crossed cylinders." *Optician*, **74**, 1908, 137 (21 Oct. 1927); **74**, 1911, 203 (11 Nov. 1927); **75**, 1929, 63 (16 Mar. 1928); **75**, 1935, 169 (27 April 1928).
TOMKINS, J. A. "Crossed cylindrical lenses." *Optician*, **81**, 2097, 275 (5 June 1931).
SMITH, F. G. "Crossed cylindrical lenses in contact." *Optician*, **88**, 2272, 143 (12 Oct. 1934).
TOMKINS, J. A. "On the prismatic effect of an astigmatic lens." *Optician*, **81**, 2084, 39 (6 March 1931).
SMITH, T. "On toric lenses." *Trans. opt. Soc.*, Lond., **29**, 2, 71 (1928).
(10) WHITWELL, A. "Fused bifocal lenses." *Trans. opt. Soc.*, Lond., **32**, 3, 101 (1931).
WHITWELL, A. "Lenticular cataract lenses." *Optician*, **83**, 2144, 174 (29 April 1932).
WHITWELL, A. "On the best form of spectacle lens. Presbyopic correction." *Optician*, **89**, 2301, 203 (3 May 1935).
(11) *Contact Glasses.* Zeiss Leaflet with Bibliography; FEINBLOOM, W. *Amer. J. Optom.* **9**, 3, 78 (1932).
(12) *Telescopic Spectacles.* Zeiss Leaflet with Bibliography.

Refraction

(13) FREEMAN, H. and RIND, H. "A comparison of subjective and objective findings in the determination of ocular astigmatism." *Refractionist*, **18**, 267, 190 (May 16 1929).
SWAINE, W. "A comparison of refractive tests and the latent aberration interval." *Optician*, **78**, 2013, 161 (25 Oct. 1929); **78**, 2014, 187 (1 Nov. 1929).
KURTZ, J. I. "A comparative study of refraction with and without cycloplegia." *Amer. J. Optom.* **11**, 7, 248 (1934).
KUHL, A. "Accuracy of refraction measurements of the eye." *Z. tech. Phys.* **14**, 7, 294 (1933).
GJESSING. "Comparative estimations of refraction." *Acta ophthal.*, Kbh., **6**, 222 (1928).
(14) BENTALL, W. K. and DIPROSE, D. R. "A practical examination of certain theoretical aspects and anomalies connected with static retinoscopy." *Trans. Inst. ophthal. Opt.* **1** (1932).
(15) STINE, G. H. "Variations in refraction of visual and extra-visual pupillary zones." *Amer. J. Ophthal.* (Feb. 1930).
(16) FERREE, C. E. and RAND, G. "The refractive conditions of the peripheral field of vision." See Ref. (2), p. 244.
(17) AMES, A. and GLIDDON, G. H. "Ocular measurements." *Trans. Amer. med. Ass.*, ophthalmology (1928).
FINCHAM, E. F. "Some observations on the dynamic refraction of the eyes." *Refractionist*, **23**, 398, 350 (1 Nov. 1934).
FERREE, C. E., RAND, G. and LEWIS, E. F. "The effect of intensity of light on visual acuity of presbyopic and non-presbyopic eyes." *Trans. Illum. Eng. Soc. Amer.* **29**, 296 (1934).
(18) AMES, A., GLIDDON, G. H. and OGLE, K. N. "Size and shape of ocular images." *Arch. Ophthal.* **7**, 576 (1932).
AMES, A. and OGLE, K. N. *Arch. Ophthal.* p. 904.
CARLETON, E. H. and MADIGAN, L. F. *Arch. Ophthal.* p. 720.
AMES, A., OGLE, K. N. and GLIDDON, G. H. "Corresponding retinal points, the horopter and size and shape of ocular images." *J. opt. Soc. Amer.* **22**, 10 and 11, 538 (1932).
OGLE, K. N. "Analytical treatment of the longitudinal horopter." *J. opt. Soc. Amer.* **22**, 12, 666 (1932).
DOANE, H. C. "Clinical significance of differences in relative size and shape of ocular images." *Amer. J. Optom.* **11**, 10, 390 (1934).
DOANE, H. C. "Clinical developments in correction of aniseikonia." *Amer. J. Optom.* **13**, 2, 42 (1936).

BIELSCHOWSKY, A. "Congenital and acquired deficiencies of fusion." *Amer. J. Ophthal.* **18**, 10 (1935).

(19) PASCAL, J. I. *Modern Retinoscopy* (1930).

(20) COATES, W. R. "Visual acuity and test letters." *Trans. Inst. ophthal. Opt.* **3** (1935).
EMSLEY, H. H. "Visual acuity and the I.O.O. Test Chart." *Trans. Inst. ophthal. Opt.* **2** (1933).
ADAMSON, J. "Visual acuity tests." *Refractionist*, **22**, 376, 406 (1 Dec. 1933).
WOODRUFF, E. W. "Some factors affecting visual acuity." *Optician*, **82**, 2116, 147 (1931).
BANISTER, H. "On the selection of tests of visual acuity." *Refractionist*, **20**, 316, 214 (1 June 1931); **20**, 319, 275 (16 July 1931).

(21) LUCKIESH, M. and MOSS, F. K. *Seeing* (1931).
LYTHGOE, R. J. "The measurement of visual acuity." *Med. Res. Coun.* (1932).

(22) HAMBLIN, J. R. and WINSER, T. H. "On the resolution of gratings by the astigmatic eye." *Trans. opt. Soc.* **29**, 10, 28 (1927).

(23) See papers by CREED, R. S., HARTRIDGE, H. (Ref. 2).
HOUSTOUN, R. A. *Vision and Colour Vision* (1932).
MARTIN, L. C. *Proc. Opt. Convention*, p. 166 (1926).

(24) FREEMAN, E. "Anomalies of visual acuity in relation to stimulus distance." *J. opt. Soc. Amer.* **22**, 5, 285 (1932).
LUCKIESH, M. and MOSS, F. K. "Dependency of visual acuity upon stimulus distance." *J. opt. Soc. Amer.* **23**, 1, 25 (1933).

(25) EMSLEY, H. H. "The effect of visual acuity on the recognition of moving objects in peripheral parts of the field of vision." *Interim Report Inst. ophthal. Opt.* (1928).
GRENDLEY, G. C. "Psychological factors in peripheral vision." *Med. Res. Counc.* 163 (1932).

(26) WESTON, H. C. and ADAMS, S. "Further experiments on the use of special spectacles in very fine processes." *Med. Res. Coun.* (1929).
DUKE-ELDER, W. S. "Investigation into effect on eyes of occupations involving close work." *Brit. J. Ophthal.* (Dec. 1930).
SOURASKY, A. "Race, sex and environment in development of myopia." *Brit. J. Ophthal.* **12**, 4 (1928).
EASTMAN KODAK CO. "Industrial Survey." *Opt. J. and Rev.* **66**, 22 (1932).
SWAINE, W. "Refractive errors in school children." *Interim Report Inst. ophthal. Opt.* (1928).
Report of Board of Education Committee. Defective Vision in School Children (1931).

(27) CANTONNET, A. and FILLIOZAT, J. *Strabismus* (trans. by Dr M. Coque) (1934).

(28) RAXWORTHY, A. W. S. "Relative values of the various instruments used in dealing with cases of errors in binocular vision." *Refractionist*, **18**, 275, 338 (16 Sept. 1929).

(29) AMERICAN OPTICAL COMPANY. *Rhythm Reading with the Ophthalmograph and Metronoscope.* Compare VERNON, M. D. "The movements of the eyes in reading." *Med. Res. Coun.* (1930).
ADLER, F. H. and FLIEGELMAN, M. "Influence of fixation on visual acuity." *Arch. Ophthal.* (Oct. 1934).

(30) See papers by FINCHAM, E. F., PFLUGK, A. v. (Ref. 2).

(31) FINCHAM, E. F. "A study of accommodation by photography of the living lens and ciliary body in a case of Aniridia." *Trans. ophthal. Soc.* **55** (1935).

(32) ADAMSON, J. "A study of the cyclo-rotational powers of the eyes." *Trans. opt. Soc.* **33**, 5, 189 (1932).

Ophthalmic instruments

Astigmatic Charts

(33) WHITWELL, A. "V in Astigmatic Charts." *Optician*, **75**, 1950, 451 (10 Aug. 1928).
WOODRUFF, E. W. *Optician*, **76**, 1952, 3 (24 Aug. 1928).
WHITWELL, A. *Optician*, **76**, 1962, 183 (2 Nov. 1928).
PASCAL, J. I. "Ioskiascopy test (Raubitschek)." *Arch. Ophthal.* **7**, 378 (1932).
EMSLEY, H. H. "Fan charts for subjective testing of astigmatism." *Interim Report Inst. ophthal. Opt.* (1928).

Visual acuity

(34) *Clason Acuity Meter.* Bausch and Lomb Catalogue.
FERREE, C. E. and RAND, G. *Projector.* Bausch and Lomb Catalogue.

Subjective refraction

(35) BROWN, C. "Duochrome refraction test." *Trans. opt. Soc.* **31**, 1, 50 (1929).
ROESSLER, D. F. "Subjective test for astigmatism." (Ref. 2).
WHITWELL, A. "Scheiner Disc." *Optician*, **77**, 2000, 417 (26 July 1929).
LINDNER, K. "Velono Streak test." *Z. Augenheilk.* **60**, 346 (1926). Compare JACKSON, E. *Optician*, **83**, 2155, 355 (15 July 1932).

Objective refraction

(36) WHITWELL, A. "Busch optometer." *Optician*, **76**, 1973, 380 (18 Jan. 1929).
FINCHAM, E. F. "Coincidence optometer." Twenty-fifth exhibition of instruments. Physical Society (1935).
WHITWELL, A. "Fincham coincidence optometer." *Optician*, **90**, 2339, 405 (24 Jan. 1936).

§ 6. COLOUR-VISION

By W. D. WRIGHT, D.Sc., Ph.D., A.R.C.S., D.I.C.

The problems of colour-vision are being attacked by both physicists and physiologists along four broad fronts: (*a*) by the measurement of the colour vision characteristics of different types of observer; (*b*) by various psycho-physical experiments; (*c*) by examination of the structure and the chemistry of the retina; (*d*) by the measurement of retinal action potentials and the recording of nerve-impulses.

Heading (*a*) concerns the measurement of those characteristics of an observer's eye which determine the properties of his colour perception mechanism. These characteristics include the colour mixture functions and the hue and saturation discrimination data. The information regarding normal trichromats is now fairly complete, but there have been some wide gaps in our knowledge of the so-called "colour-blind" observers. Two papers published in 1935, one by Pitt[1] and the other by Trendelenberg and Schmidt[2], have helped considerably to fill in the gaps. Pitt's work was largely confined to dichromats, and his report provides very complete information for a relatively small number of observers. The luminosity curves, dichromatic coefficient curves, and hue discrimination curves are given for six protanopes and six deuteranopes. The two groups appear to be well defined and comparatively easy to distinguish from the trichromat by their reduced hue discrimination in the green-red part of the spectrum, and from one another by having very different luminosity curves. The data obtained have been used to subdivide the trichromatic colour triangle into a series of isochromatic zones, where any zone represents the locus of colours that, although seen by the trichromat as quite distinct, are indistinguishable to the dichromat when adjusted to the same intensity level. Introductory data on 3 anomalous trichromats are also given. The paper by Trendelenberg and Schmidt is in a sense complementary to that of Pitt, in that they made a simple test with the Nagel anomaloscope on a large number of observers, yielding 129 colour defective persons, and subdivided the latter into the four broad

groups of protanopes, deuteranopes, protanomalies and deuteranomalies. In addition to the relative distribution of the four groups, the important fact emerges that the two groups of anomalous trichromats are well-defined among themselves and do not grade into either the normal trichromat on the one side nor into the dichromat on the other. This is of very great theoretical interest as it appears to cut across the idea of any gradual evolution of colour vision, for on this basis a continuous distribution of observers from the monochromat through the dichromat to the trichromat, would be expected. The correct explanation must for the present remain a matter of speculation.

Section (*b*) includes all variety of experiments on flicker, adaptation, contrast, interaction between retinal areas, and so on. In a paper by Graham and Marajaria[3], curves are given showing the relation between the intensity of a patch of light and the time of exposure for the patch to be just visible, when viewed in the peripheral retina. As the area of the field is reduced, the curves tend to approximate to similar data obtained by measuring the electrical response from a single cell in the eye of *limulus*. It is concluded that the statistical distribution of the characteristics of a number of cells is responsible for the experimental curves and for their change with area. This conclusion is interesting, among other reasons, as an illustration of the part that statistical variations of sensitivity and response must play in visual perception, where large numbers of nerve endings are acting simultaneously.

Hecht and his collaborators have obtained results on the critical frequency of flicker and its variation with field size[4], and the results provide still further evidence of the separate activities of rods and cones. This group of workers have also been responsible for experimental and theoretical work on adaptation and intensity discrimination[5,6], but experiments described by the present writer[7], do not support Hecht's photo-chemical theory of intensity discrimination. These experiments demonstrate that there is no direct connexion between the concentration of the photo-chemical substance in the retina and the increase of a stimulus necessary to produce a just noticeable intensity difference.

Among other contributions that should be included under heading (*b*), some very suggestive work by Bouman[8] may be mentioned. Results on the electrical excitability of the eye have led to the conclusion that the origin of the excitation is not in the rods or cones, but in the nerve fibres either in the retina or in the optic nerve. He has found no evidence for two thresholds, and the level of excitability is determined by the state of adaptation. Flicker phenomena were observed and gave a fusion rate of 120 per sec.

Although little has been published during 1935 that might be included under heading (*c*), Tansley[9] has reported differences in the chemical constitution of the light and dark-adapted eye of both the frog and the rat, while Hosoya and Saito[10] have confirmed the production of visual yellow from visual purple, and claim to have produced a visual white which they regard as a very pure visual yellow. It is obvious that all information about the constitution of the light-sensitive substances in the retina must have a very direct bearing on our ideas of the visual processes, but at the same time the difficulties of isolating sufficiently pure material in any but very

small quantities, make this method of approach to the problem arduous, though attractive.

Some exhaustive work by Österberg[11] on the topography of the human retina, has added considerably to our knowledge of the relative distribution of the rods and cones. He found, for example, that the number of rods increases to a maximum at a distance of about 5–6 mm. from the fovea, and that the upper and nasal parts of the retina are better supplied with rods than the lower and temporal parts. The total number of receptors in the human retina was found to be between 110,000,000 and 125,000,000 rods and between 6,300,000 and 6,800,000 cones.

The work under heading (*d*) on retinal action potentials has been, perhaps, the most fruitful of any. The researches of Granit and his colleagues[12, 13], have resulted in a further analysis of the different components of the potential. The relation between the components for different conditions of adaptation has been examined and the increase of inhibition with light adaptation established. Two types of retinae have been found and named *E*-retinae and *I*-retinae, which are characterized by several distinctive features, of which the chief is the direction of the response when subject to intermittent stimuli. There is the possibility that *E*-retinae, of which the cat's is an example, may correspond to rod-retinae, while *I*-retinae, as exemplified by the pigeon, may correspond to cone-retinae.

Graham and Hartline[14] have recorded the responses of single cells of *limulus*, but have found no variation in the type of response as the wave-length of the light stimulus was varied. Small variations in the luminosity curves of different nerve endings were, however, detected.

Graham, Kemp and Riggs[15] have experimented with a pigeon's eye (a colour-discriminating eye) and have recorded the responses, using both white and coloured stimuli. They found that the type of response for one wave-length could always be matched by the response for any other wave-length, when a suitable intensity was used. This result is in disagreement with similar work by Kohlrausch and Brossa[16], and no final conclusions can be arrived at yet, especially as the magnitude of the responses seems to have been rather small. The question is, however, of vital importance since, if any specific effect were found, it would be the first direct clue to the individual character of the different colour responses. It is to be hoped that further work on these lines will be carried out.

Three or four papers that belong to a section on colorimetry rather than colour vision, but which may for convenience be discussed here, include one by Judd[17] who has derived a colorimetric co-ordinate system by trial and error, in which the length of any line on the Maxwell triangle is a close representation of the magnitude of the subjective colour difference as seen by the normal observer. Two valuable papers by MacAdam[18, 19], give the results of calculations by which the spectrophotometric curve yielding the maximum luminosity for any given point in the colour triangle can be determined. And finally a new trichromatic colorimeter in which the mixing is carried out in an integrating sphere, has been designed by Donaldson[20] and combines simplicity of operation with accuracy of result.

In this review reference has been made to papers that seemed to the writer to

be typical of the work in progress and is in no sense an exhaustive survey. The bibliography includes only those papers directly referred to, and as these have been selected from a list of over 80 papers published in 1935, it will be appreciated that much work of value has had to pass unnoticed.

At least one point should emerge from this review, namely the necessity for the close co-operation of physicists and physiologists, and the importance of each group of workers keeping in touch with the progress made in other fields.

References

(1) Pitt, F. H. G. *Spe. Rep. Ser. Med. Res. Coun.*, London, No. 200 (1935).
(2) Trendelenberg, W. and Schmidt, J. *S.B.* Preuss. Akad. Wiss. **2** (1935).
(3) Graham, C. H. and Marajaria, R. *Amer. J. Physiol.* **113**, 299 (1935).
(4) Hecht, S. Schlaer, S. and Smith, E. L. *Quantitative Biology*, **3**, 237 (1935).
(5) Hecht, S., Haig, C. and Wald, G. *J. gen. Physiol.* **19**, 321 (1935).
(6) Hecht, S. *J. gen. Physiol.* **18**, 767 (1935).
(7) Wright, W. D. *J. Physiol.* **83**, 466 (1935).
(8) Bouman, H. D. *Arch néerl. Physiol.* **20**, 430 (1935).
(9) Tansley, K. *J. Physiol.* **84**, 34 *P* (1935).
(10) Hosoya, Y. and Saito, Z. *Tohôku J. exp. Med.* **27**, 172 (1935).
(11) Österberg, G. Thesis Copenhagen Univ., Nyt Nordisk Forlag, Copenhagen (1935).
(12) Granit, R. and Therman, P. O. *J. Physiol.* **83**, 359 (1935).
(13) Granit, R. *J. Physiol.* **85**, 421 (1935).
(14) Graham, C. H. and Hartline, H. K. *J. gen. Physiol.* **18**, 917 (1935).
(15) Graham, C. H., Kemp, E. H. and Riggs, L. A. *J. gen. Psychol.* **13**, 275 (1935).
(16) Kohlrausch, A. and Brossa, A. *Arch. Physiol.* **421** (1914).
(17) Judd, D. B. *Bur. Stand. J. Res.*, Wash., **14**, 41 (1935).
(18) MacAdam, D. L. *J. opt. Soc. Amer.* **25**, 249 (1935).
(19) MacAdam, D. L. *J. opt. Soc. Amer.* **25**, 361 (1935).
(20) Donaldson, R. *Proc. phys. Soc.* **47**, 1068 (1935).

§7. GLARE

By W. S. Stiles, Ph.D.

The terms glare and dazzle are used to denote the effects which occur when the ease or efficacy of vision is impaired as the result of a high or misplaced brightness. The effects may be transient or may persist so long as a particular disposition of brightness is present in the field of view. Common examples of glare are the discomfort experienced when the eyes are exposed to the high brightness of white pavements in sunlight and the temporary blinding produced by motor car headlights. With the development of illuminating engineering in the last thirty years, considerable attention has been paid to the subject of glare mainly with the object of discovering how artificial lighting of all kinds can be designed so that glare is eliminated or minimized. The results of many of the earlier investigations have been summarized by Lythgoe[1] and by Millar and McGray[2]. These results tend to be scattered observations and isolated sets of data rather than a connected body of knowledge, partly because of the complexity of the glare problem and partly because of differences in the aims and standpoint of the investigators who include psycho-

logists, physiologists, physicists and illuminating engineers. In what follows no attempt has been made to condense into a few pages existing information on all aspects of the glare problem. Instead an outline is given of the progress made in studies dealing with certain glare phenomena which belong to physiological optics. As a preliminary however it may be useful to describe briefly the classification of glare phenomena which was recommended to the attention of research workers at the 1935 meeting of the International Commission on Illumination[3].

The basis of the classification is set forth in tables 1 and 1A. In the first main column of table 1 the various ways in which glare may affect the individual are classed under five main headings labelled A to E. The headings are self-explanatory except possibly C which is meant to include pupillary contraction, blinking, watering of the eyes and so on. Different types of brightness distribution which may produce glare effects are specified in the second main column, the first sub-column giving the external brightness conditions, the second, the subject's direction of view. The two together determine the effective brightness distribution to which the eyes are subjected, with the limitation that blinking and involuntary eye motions cause abrupt intermissions and translations of the brightness distribution which can hardly be avoided even in controlled laboratory experiments. The final column defines the various ways in which the existence of a glare effect may be established or brought to notice. It is necessary to include this factor as with our present knowledge it is seldom possible to say whether effects observed in different ways have a common cause. A given case of glare is fitted in the classification by indicating the headings under which it falls in each of the columns, e.g. D-1·4*c*-I, B-1·2*a*-III. In many glare phenomena a comparison is implied between a brightness distribution in which the glare effect is experienced, and a brightness distribution differing in some particular but otherwise resembling the first, in which the glare effect does not occur. In such cases it is necessary to specify the comparison brightness distribution (see table 1A). Effects on the eye or on vision due to invisible radiation (ultra-violet or infra-red) are not included in the classification.

Effect of glare on visual acuity. The classification just described tells us little about the actual phenomena of glare but it does provide a useful method of sorting out the different effects into groups meriting separate study. Among these groups the one on which most work has been done is specified as B-1·4*a*-III in the classification and, in the simplest case, corresponds to the following conditions of observation.

An observer directs his gaze at a test object by means of which a numerical measure of visual performance can be made. At a given point in the field of view is an exposed light source or a small area of relatively high brightness which forms the glare source, and the experiment consists in determining the change in visual performance due to the presence of the glare source and the dependence of this change on such factors as the brightness of the test object and the intensity and position of the glare source.

In the first measurements on these lines visual acuity was used as the test of visual performance. The following conclusions drawn by Cobb[4] from his own experiments are consistent with most of the earlier observations. (1) The vision of

Table 1. *Classification of glare phenomena*

Type of effect produced	Type of brightness distribution producing the effect	Direction of view	Evidence for the effect
A Sensation of discomfort, distraction or strain B Measurable reduction in visual capacities or in the efficiency of performance of work involving vision in some degree C Modifications in the involuntary physiological reaction to the situation D After-image and irradiation effects E Pathological after-effects	1 Static distributions 1·1 High general level of brightness or high brightnesses in the working area 1·2 Working area brighter than surround 1·3 Working area darker than surround 1·4 Field containing exposed light sources or patches of high brightness other than (possibly) the working area 1·5 Field containing a more or less uniform brightness superposed on the normal brightnesses of the working area 2 Distributions varying periodically 2·1 Rapid flicker of all brightnesses in the field 2·2 Relatively slow cyclic changes in the brightness distribution 3 Variable but non-periodic distributions 3·1 Abrupt change in brightness level 3·2 Abrupt change in brightness, area or position of exposed light sources 3·3 Gradual changes in the brightness level or in the brightness area or position of exposed light sources	*a* Steady fixation *a*. 1 Fixation point within working area *a*. 2 Fixation point exterior to working area *b* Direction of vision confined to a small region of the field *c* Direction of vision wandering without restraint over the whole or a large part of the field	I Casual observation II Personal opinions of one or more observers in organized tests under controlled conditions III Direct laboratory measurements under completely controlled conditions IV Quantitative experiments under practical conditions (e.g. studies of output in factories or driving speeds on roads in relation to illumination) V Statistics correlating lighting with industrial output, accident rate, etc. VI Clinical observation and case histories

N.B. By "working area" is meant that part of the field containing the brightness differences which must be perceived in carrying out a particular visual test or any task involving vision.

an object is lowered progressively as the glare source becomes brighter and the angle formed by it and the line of vision decreases. (2) If the test object is very bright the presence of a weak glare source may cause increased visual acuity. (3) When the glare source is 10° from the line of vision and the lateral illumination of the eye and the illumination of the test object are equal, there is no appreciable reduction of acuity at any intensity. (4) The presence of the retinal image of the glare source is a negligible factor in the depression of vision; diffusion and reflection

Table 1A. *Practical examples of the various types of brightness distribution included in the glare classification* (*table* 1), *together with the comparison brightness distribution for each case*

1·1.	White roads or pavements in full sunlight. Vision may be less good than if all brightnesses in the field were reduced in the same proportion
1·2.	Local lighting of work in the absence of adequate general room lighting. Vision may be less good than if general lighting were added so that the brightnesses of working area and surround were approximately the same
1·3.	High general room lighting with some local lighting of low reflection factor working material. Vision may be less good than if the general lighting were reduced so that the brightnesses of working area and surround were approximately the same
1·4.	Street illumination. Vision may be less good than if the light sources were screened from view leaving the brightness distribution otherwise unchanged
1·5.	Specular reflexion of a window or a light source from the printed page. Vision may be less good than if the superposed brightness of the specular reflexion were eliminated leaving the normal diffuse brightness unchanged
2·1.	Artificial electric lighting using low frequency alternating current. Vision may be less good than if the flickering light sources were replaced by steady sources giving at each point of the field the same average brightness
2·2.	Artificially lighted street as seen by the driver of a moving car. Vision at any instant may be less good than if the changing brightness distribution were replaced by the steady brightness distribution obtained by stopping the car at the instant in question
3·1.	Lighted room as seen by an observer entering from the dark. Immediately after the abrupt change vision may be less good than when equilibrium with the final condition has been reached
3·2.	Flash of the headlights of a turning car when the beams sweep across the eyes of an observer. Immediately after the flash vision may be less good than if the flash had not occurred
3·3.	Appearance, transit of the field and disappearance of the headlights of an approaching car as seen by a driver at night. In this succession of events vision at any instant may be less good than if the changing brightness distribution were replaced by the steady brightness distribution obtained by stopping the cars at the instant in question

from the eye media are the important factors, since the reduction of acuity still occurs when the image of the glare source is formed on the blind spot. (5) The effects of a glare source can be imitated by throwing a haze of light over the central retina and test object.

Recent work by Lythgoe[5] has confirmed the second of the conclusions stated above and has further revealed that in most cases an improvement in acuity due to the presence of a glare source is accompanied by a contraction of the pupil of the eye and is eliminated by the use of artificial pupils. Lythgoe regards the effect of a glare source on visual acuity as a resultant effect of an increase in acuity due to pupillary contraction with improvement in the optical system of the eye and a decrease due to an effect of retinal origin or due to scattering of light within the eye.

It should be noted that Lythgoe's measurements were confined to feeble glare sources producing changes of acuity of less than 10 per cent. The glare was therefore much less than that encountered in the motor car headlight problem. Born and Wolff[6] studied glare intensities in the headlight range but their work is connected more closely with the investigations discussed in the next section.

Effect of a glare source on the brightness-difference threshold. The effect of a glare source on the smallest perceptible difference of brightness T between the two halves of a photometric field or between a simple test object and the background against which it is seen, has been studied in recent years particularly by Bordoni, Luckiesh and Holladay, Stiles and Crawford[7–14]. It will be recalled that, according to Weber's Law, in the absence of glare the threshold difference T is proportional to the brightness B of the photometric field or of the background, i.e. $(T/B) = \text{constant}$. However this result is only an approximation valid for a certain range of moderate and high brightnesses B and must be replaced by $(T/B) = f(B)$ where $f(B)$ increases with decrease of B, slowly at first and then rapidly when B falls below about 0·01 candles/sq. ft. The effect of a glare source introduced into the field without changing B is always to increase T, and the following quantitative results have been established:

(1) If the angular diameter of the glare source is not too great, the increase in T due to a glare source in a fixed position is determined by the illumination E which it produces at the eyes of the observer and is independent of its brightness and angular dimensions taken separately[7, 8, 9].

(2) For a given angle θ between the direction of vision and the glare source the increase in T is independent of the azimuth in which the glare source lies[8, 9, 12].

(3) The increase in T due to the glare source is equal to the increase which would be produced in the absence of glare by increasing the field brightness from B to β where β is given by the formula:

$$\beta = B + kE/\theta^n, \qquad \text{.....(1)}$$

k and n being constants[8, 9, 11, 12].

β is termed the equivalent background brightness or equivalent uniform field brightness.

(4) The value of β for a glare source in a given position and of given intensity is approximately the same whatever the size, exposure time or colour of the brightness difference used in the determinations of T on which the derivation of β is based. This is provided the angular diameter of the test object is small compared with θ[13].

(5) If there are two or more glare sources in the field of view at angles θ_1, θ_2, etc. with the direction of vision and producing eye illuminations E_1, E_2, etc., then the equivalent uniform field brightness β has the value:

$$\beta = B + k\Sigma\left\{\frac{E_1}{\theta_1^n} + \frac{E_2}{\theta_2^n} + \ldots\right\} \qquad \text{.....(2)}^{(8,9)}.$$

Table 2 gives the values of the constants k and n obtained in the several investigations. The value of k depends on the units in which θ, B, E and β are expressed, which have here been taken as candles per square foot for β and B, foot-candles for E and degrees for θ. The measure of agreement of the values of k and n given in table 2

Table 2

Investigators	n	k	Remarks
Luckiesh and Holladay	2	13	
Holladay	2	9·2	
Stiles	1·8	7·0	Revised values allowing for the large size of test object used
Stiles and Crawford	2·1	11	Mean results for a series of determinations using a special instrument for measuring the equivalent uniform field brightness
Mean	2·0	10	

is satisfactory for work of this kind. The range of the variables B, E and θ over which formula (1) was tested varied in the different investigations but considering all the results it appears that the formula certainly holds as an approximation for $B < 0{\cdot}3$, $E < 1$ and θ between 1 and 25°. A further check on the formula is provided by some measurements of Bartley and Fry [15] at a glare angle of 5°. They found that a glare source of brightness 2220 can./ft.^{2} of visual angle 1° 14′ increased T by the same amount as an additional field brightness of 0·34 can./ft.^{2} Calculation from formula (1) with $k = 10$ and $n = 2$ gives for this case $\beta = 0{\cdot}32$ can./ ft.^{2}

In so far as formula (1) is valid, the region of the retina on which the test object is imaged (the fovea) behaves, in the presence of a glare source, *as if* on and around it an additional illumination had been imposed of an intensity corresponding to an external brightness β. It is easily verified that a small bright light in an otherwise dark field appears as the centre of a system of haloes and diffuse luminosity suggesting the presence of a distribution of scattered or diffracted light round about the retinal image of the source. Adopting this view we base the simplest explanation of the action of a glare source on the difference threshold on the following assumptions: (*a*) the visual sensitivity is determined solely by the retinal illumination in the region occupied by the image of the test object, (*b*) the rays from the glare source produce in their passage through the eye an illumination of the fovea by scattering, diffraction or irregular reflection and refraction. This theory accounts for the fact that a glare source imaged on the blind spot produces about the same increase in T as a glare source imaged at any other position equally distant from the fovea. Assuming scattering in the optic media of the eye by particles small compared with the wave-length of light, Holladay [9] derived the intensity of scattered light at the fovea from Rayleigh's law and showed that it was approximately proportional to E/θ. Compared with the experimental law for β, this result gives the correct variation with E but too slow a variation with θ. It has since been shown, however [16], that the light losses in the optic media are insufficient to account for the amount of scattering material required on this assumption. From observation of the angular diameters and intensities of the haloes seen round a bright light, Raman [17] deduced the presence in the optic media of structures averaging in size about 13 and 7 μ respectively. Structures of this size would not obey Rayleigh's law and would scatter relatively more light in the forward direction. The modifications required

in the theory if larger particles are assumed should bring it in better agreement with the experimental law. On the other hand, Pokrowski[18] has suggested that light scattered in the retina itself at the image of the glare source produces the necessary illumination of the fovea. He deduces that for a glare source of constant intensity at different angles θ the illumination of the fovea should be proportional to $E/(\theta^2+c)$, where the constant c equals about 0·5, and finds that his measurements of T over the range $\theta = 0{\cdot}6°$ to $3{\cdot}6°$ are consistent with this expression. Schouten[19] has found that for large angles of glare (25° or more) about half the effect of the glare source is due to light which penetrates the sclerotic so that the effect is much reduced at these angles if the illumination of the eye is confined to the cornea. The scleral component of glare is certainly to be ascribed to an actual scattered or diffused illumination of the retina. It plays little part at small angles of glare.

If assumption (*a*) of the scattering theories of glare be rejected as untrue even as an approximation, the action of a glare source may be accounted for in whole or in part as the result of nervous interaction in the retina, optic nerve or brain. Various possibilities of this kind are enumerated by Holladay[9] but no quantitative theory has yet been put forward.

Effect of a glare source on the difference threshold for extrafoveal vision. If the test object is imaged at a point of the retina other than the fovea, the relation between the difference threshold T and the field brightness B is more complicated than for foveal vision owing to the operation of cone vision at high brightnesses and rod vision at low brightnesses. The only published measurements of the action of a glare source on the threshold for extrafoveal vision refer to a test object viewed at 5° (external angle) from the fovea and to a glare source at 6° from the test object[13, 20]. For this case, it has been found that the equivalent uniform field brightness, defined as for foveal vision, is independent of the character of the test object used in measuring T, is proportional to the eye illumination E ($B = O$) and is of the same order of magnitude as for foveal vision. The additivity of the effect of several glare sources has been established for 5° extrafoveal vision by Crawford[21]. If formula (1) above turns out to be generally valid for extrafoveal vision, an explanation of the formula in terms of a scattering theory will become more probable.

Recovery of the eye after exposure to glare. No adequate measurements have been published on the recovery of the eye after the abrupt removal of a glare source from the field of view. Such measurements would provide a necessary test of scattering theories of glare which require the recovery curve to be identical with that following the abrupt removal of a uniform field brightness of intensity $(\beta - B)$.

Colour of the glare source. Several investigators have studied the effect of glare sources of different colours on the perception of test objects illuminated in white light[22, 6, 20]. Only small differences, if any, were observed in the effects produced by white and coloured glare sources. This refers to direct (foveal) vision of the test object. For extrafoveal vision, a feeble blue light raises the threshold more than a red light of equal intensity[20]. The glare effects of white and yellow light have been compared for the case when both glare source and object emit light of the same colour by Arndt[23] using monochromatic Na light and le Grand[24] using yellow

light obtained from the white light by interposition of a glass filter at the eyes. Neither observed any difference in the glare effect produced by white and yellow light. On the other hand, Monnier and Mouton [25] and Escher-Desrivières, Faillie and Jonnard [26] found that after staring at a very bright light source of white or yellow light the eye recovered more quickly in the case of the yellow light. Unfortunately, only very scanty details of these experiments have been published.

References

(1) Lythgoe. *Spec. Rep. Ser. Med. Res. Coun.*, Lond., No. 104 (H.M.S.O. 1926).
(2) Millar and McGray. *Proc. Int. Comm. Illum.* (1928).
(3) Sec. Rep. on Glare, *Proc. Int. Comm. Illum.* (1935).
(4) Cobb. *Amer. J. Physiol.* **29**, 76 (1911).
(5) Lythgoe. *Spec. Rep. Ser. Med. Res. Coun.*, Lond., No. 173 (H.M.S.O. 1932).
(6) Born and Wolff. *Licht*, **2** (1932).
(7) Bordoni. *Elettrotecnica*, p. 585 (1924).
(8) Luckiesh and Holladay. *Trans. Amer. Illum. Engng Soc.* **20**, 221 (1925).
(9) Holladay. *J. opt. Soc. Amer.* **12**, 271 (1926).
(10) Holladay. *J. opt. Soc. Amer.* **14**, 1 (1927).
(11) Stiles. *Proc. roy. Soc.* B, **104**, 322 (1929).
(12) Stiles. *Tech. Pap. Illumin. Res.*, Lond., **8** (H.M.S.O. 1929).
(13) Stiles and Crawford. *Phys. Soc. Discussion on Vision*, p. 194 (1932).
(14) Crawford and Stiles. *J. Sci. Instrum.* **12**, 177 (1935).
(15) Bartley and Fry. *J. opt. Soc. Amer.* **24**, 342 (1934).
(16) Stiles. *Proc. roy. Soc.* B **105**, 131 (1929).
(17) Raman. *Phil. Mag.* **38**, 568 (1919).
(18) Pokrowski. *Z. Phys.* **35**, 776 (1925).
(19) Schouten. *Proc. K. Akad. Wet. Amst.* **37**, 506 (1934).
(20) Stiles and Crawford. *Proc. roy. Soc.* B **116**, 55 (1934).
(21) Crawford. *Phys. Soc.* **48**, 35 (1936).
(22) Luckiesh, Taylor and Holladay. *J. opt. Soc. Amer.* **11**, 311 (1925).
(23) Arndt. *Licht*, **3**, 213 (1933).
(24) Le Grand. *C.R. Acad. Sci.*, Paris, **198**, 1075 (1934).
(25) Monnier and Mouton. *C.R. Acad. Sci.*, Paris, **196**, 1659 (1933).
(26) Escher-Desrivières, Faillie and Jonnard. *C.R. Acad. Sci.*, Paris, **198**, 1325 (1934).

§8. NEW LIGHT-SOURCES: RECENT DEVELOPMENTS IN ELECTRIC DISCHARGE LAMPS

By J. T. Randall, M.Sc., Member of the Staff of the Research Laboratories of the G.E.C. Ltd., Wembley, England

The recent rise of importance of the discharge lamp is a result of the search for new and more efficient sources of light. In this respect the search has been one with a commercial object in view. The result, as in many former instances, is that many interesting and extremely valuable light-sources have been found and utilized, long before their physical properties and mechanism of operation have been worked out in detail. On many points fundamental information is, in comparison with technical information, not nearly so great as could be wished; the modern discharge lamp has now completely burst the earlier shackles of confinement, and it may be some time before fundamental research can completely overtake the new developments.

Although it is unlikely that the tungsten filament lamp will cease to be the main vehicle of domestic lighting for many years to come, the use of the term "discharge *lamp*" is perhaps significant in that manufacturers are tending to make the electric discharge a substitute, in widening fields, for the incandescent filament lamp. This implies a comparable or better light output for the discharge lamp, and a reasonable colour-rendering. With the discharge lamps of the present time there is always present the problem of "auxiliary gear". The negative current—voltage characteristic of an electric discharge necessitates always the use of a current-limiting device in series with it. The advantages of the various discharge lamps for particular lighting problems will be apparent from the descriptions which follow, but it is perhaps not out of place to mention this important limitation at the outset.

Types of discharge lamp. For the purposes of this article, discharge tubes may be divided into the five classes indicated in the following table. The object of this table is to give at a glance the main features of each type, and its normal usage. For this purpose, certain guiding figures, such as current, circuit voltage, and efficiency, have been inserted. It will be realized that in a summary of this nature the averages may differ appreciably from the figures for an individual tube or lamp within the knowledge of the reader. This applies particularly to the cold cathode section where many currents and different lengths of tubing are used to produce required effects. Figures for the remaining individual types may be found under appropriate headings later in the article. In the case of the laboratory lamps which are used primarily as spectroscopic standards for excitation or other purposes, brightness of the source, rather than efficiency, is important; and in the special section on these lamps, values of the former will be given.

In what follows, attention will be paid for the most part to new features and ideas developed during the last three or four years, but it has been found necessary in some cases to go somewhat further into the earlier development.

Cold cathode high voltage discharge tubes. The origins of this type of discharge may be traced with certainty to the eighteenth century, and an excellent summary of this aspect has recently been given[1]. For our purpose, a cold cathode discharge tube is defined as one whose electrodes are sensibly cold when the tube is in operation; that is to say, any energy dissipated by the discharge at the electrodes is not sufficient to raise them to a temperature at which thermionic emission is appreciable. The electrodes usually take the form of hollow open-ended, iron cylinders, and in consequence of the lack of thermionic emission, the cathode fall is high—of the order of a few hundred volts—and the starting voltage also correspondingly high. The currents used in tubes of this kind are low, and usually vary from about 20 to 250 mA. For tubes running at 250 mA., very large electrodes are required if the tube is to keep strictly within the cold cathode class. Higher values of current lead to overheating of the electrodes, rapid sputtering, and short life. In order to obtain currents of say 0·5 amperes, tubes are sometimes made with what is effectively a thermionic electrode, although no external heating is used. Such tubes are frequently, from the purely technical point of view, classed with cold cathode tubes, since they also usually require voltages in excess of the normal supply values to

Table 1. Guiding summary of main discharge lamp types and their uses*

Type of discharge	Excited gas or vapour	Order of total pressure (mm.)	Order of running tube current (amperes)	Circuit voltage	Approximate luminous efficiencies of tubes (lumens/watt)	Guide to effective colour	Normal uses
Cold cathode	Carbon dioxide	1·5	0·3	> 1000	2–3	Daylight	Daylight units for colour-matching
	Helium	3–4	0·035	> 1000	2	Ivory	Advertisement signs
	Nitrogen	1–2	0·3	> 1000	6	Orange-pink	Interior lighting
	Neon	5–10	0·03–0·06	Frequently > 1000	10–15	Red	Advertisement signs
	Neon-fluorescent yellow	5–10	0·03	Frequently > 1000	> 20–25	Yellow	Advertisement signs and interior lighting
	Neon-fluorescent pink	5–10	0·03	Frequently > 1000	15–20	Pink	Interior lighting and signs
	Mercury-argon	5–10	0·03	*ca.* 1000	5	Blue	Advertisement signs
	Mercury-argon fluorescent (1)	5–10	0·03	*ca.* 1000	40	Green	Signs
	Mercury-argon fluorescent (2)	5–10	0·03	*ca.* 1000	15	Blue	Signs
Hot cathode (low pressure)	Neon	2	2–10	200	11	Red	Floodlighting
	Mercury-argon	2	2–10	200	10–15	Blue or green, according to glass used	Floodlighting
	Sodium	10	2	200	40–60	Yellow	Floodlighting and road lighting
Hot cathode (low pressure laboratory types)	Mercury, cadmium-mercury, cadmium, and sodium, all generally similar to above low pressure hot cathode lamps, but much smaller in size and wattage consumed. See special section and tables 4, 5 and 6 for details.						
Hot cathode (high pressure)	Mercury-argon	1–2 atmos. (glass)	400 W.	200	45	White, deficient in red	Street lighting, industrial lighting and floodlighting
		8–10 atmos. (quartz)	80 W.	200	40		
Hot cathode (extra high pressure)	Mercury-argon	10–100 atmos. (quartz)	0·5–5 kW.		*ca.* 50	White	Projection purposes

* Except for cold cathode tubes, the reader is referred to more detailed results in the following sections.

start them. Tubes of this type are, however, rather exceptional and need not further concern us.

Everyone is familiar with the neon sign which first raised the high voltage glow discharge into prominence almost thirty years ago. As indicated in table 1, many other gases are now used, for example, nitrogen, carbon dioxide, helium, and, in addition, mercury vapour. These combinations are also common knowledge. The mercury-argon glow discharge of the cold cathode tube is pale blue in colour, and the efficiency on account of the low mercury vapour pressure is not very high. Many variations on this basic colour have been produced by the use of coloured glasses.

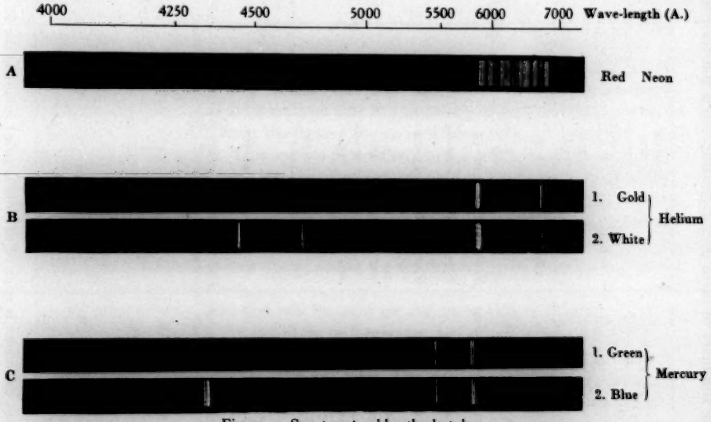


Figure 1. Spectra ot cold cathode tubes.

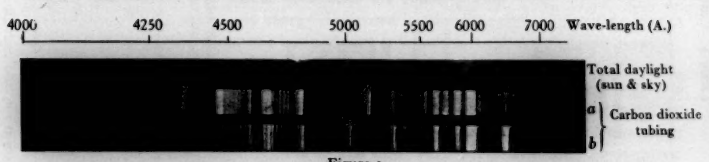


Figure 2.

By this means, numerous unusual colours may be obtained, but the luminous efficiency, already rather low, is much decreased by selective absorption. The spectrum of any discharge consists, of course, of a series of lines or bands dependent on the element or molecule excited, and the type of excitation. Thus the mercury spectrum normally obtained in a discharge tube of this type is that of the neutral

atom (HgI spectrum); in the carbon dioxide tube there are several spectral systems, one of which is the Ångström band spectrum. In figure 1 are shown the spectra of cold cathode tubes containing neon, helium, and mercury respectively. The gold helium tube and the green mercury tube are produced by the use of special glasses. Figure 2 compares the band spectrum of a carbon dioxide tube with total daylight; the apparent minimum in the green is due to the characteristic of the panchrometic plate.

The most striking developments in connexion with cold cathode tube lighting during the last two years are concerned with the use of new and highly efficient luminescent materials on the interior walls of the tubing in order to produce efficiency magnification and colour changes. A luminescent powder is, for the present purpose, one which transforms otherwise useless ultra-violet radiation into useful visible radiation. By this means at least the effective colour of the discharge is modified, if not the efficiency. For some time past powders of the zinc-cadmium sulphide type had been used to produce colour changes in the cold cathode tube mercury discharge, but these powders produced no appreciable increase in visible light output. In order to appreciate the recent changes, it is necessary to indicate a little more fully the nature and peculiarities of luminescent materials in general.

Inorganic luminescent solids are essentially exceedingly pure materials to which minute traces (frequently not more than 1 in 10^5) of some other element have not only been added but incorporated in solid-solution*. In the sulphide class it is possible to change the colour of the luminescent spectrum by either changing the nature of the impurity or by varying the relative amounts of the zinc and cadmium sulphides which form the matrix. It is a fact of experience that each class of luminescent solids responds more particularly to one band of exciting wave-lengths than another. For example, the sulphides in general respond a good deal better to long wave-length ultra-violet in the neighbourhood of 3600 A. than they do to the 2537–2000 A. range. At the same time they also luminesce very brightly under fast cathode rays. In luminous discharge tubes, the possible exciting agents are ultra-violet radiations, electrons, and, less likely, gaseous ions. It is therefore difficult to discover which of these agencies is the more important; experience suggests so far that ultra-violet radiations are at least as important as any others.

For these reasons, it is necessary to choose a powder which best responds to the radiations available. In the cold cathode mercury discharge, the 2537 A. line is extremely strong, and it is probably for this reason that the sulphides do not lead to high efficiencies in these tubes. Willemite (zinc orthosilicate) and calcium tungstate have long been known as luminescent solids, but they have not, until recently, been made with sufficient care to be of much use in discharge tubes. The new products respond very readily to the line 2537 A., and give high magnifications of luminous output. In the case of zinc orthosilicate, the magnification factor is in the neighbourhood of seven or eight to one. In figure 3, microphotometer records of the luminescent spectra of zinc orthosilicate and calcium tungstate, transformed

* There are of course other types, but the impurity phosphors are the most widely used in all technical applications.

to an approximate energy scale, are given. It is clear from these that the spectra are extremely amorphous and ill-defined. Both visually and photographically they appear as broad bands. Figure 4 shows the spectrum of a zinc-cadmium sulphide powder, and it should be noted that, in this case, the spectrum consists of two bands; this is frequently the case with sulphides, particularly when more than one activating impurity is used.

Until recently, the excitation of luminescent materials in discharge tubes had invariably been carried out with a mercury discharge. While high efficiencies were thus made possible, the effective colour of the light was not suitable for interior lighting. Rather surprisingly, it has been found that some of these powders are capable of excitation by neon radiations, and it is believed that the resonance radiation in the Schumann region is partly responsible. With neon, resultant colours

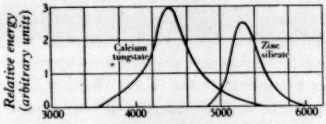


Figure 3. The distribution of energy in the luminescent spectra of zinc orthosilicate and calcium tungstate, both excited by Hg 2537.

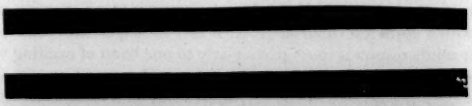


Figure 4. The upper spectrum is that of a typical zinc-cadmium sulphide luminescent powder; the lower spectrum is that of helium.

of yellow and pink, instead of the pure red, nave been obtained, and again at a considerably enhanced efficiency. It will be realized from the spectrum illustrations of figure 5 that the total light output of the tube is a combination of the discharge tube spectrum and that of a luminescent solid. The possibilities of applying certain luminescent powders to other types of discharge lamp will be mentioned later.

The development of new or more efficient colours in this way has suggested the use of this type of tubing for interior lighting in large rooms or entrance halls, to mention only two possibilities, where the use of long tubes and the inherent difficulty of high voltage are not regarded as drawbacks. The use of long tubes is inevitable in this case for the following reason. The light output from a low pressure discharge of this class is very approximately proportional to the current flowing through it; the wattage of an 8 foot mercury tube 0·6 in. in diameter is roughly 25, with a current of 0·035 amperes. Even at an efficiency of 30 lumens per watt, this means a total flux of only 750 lumens.

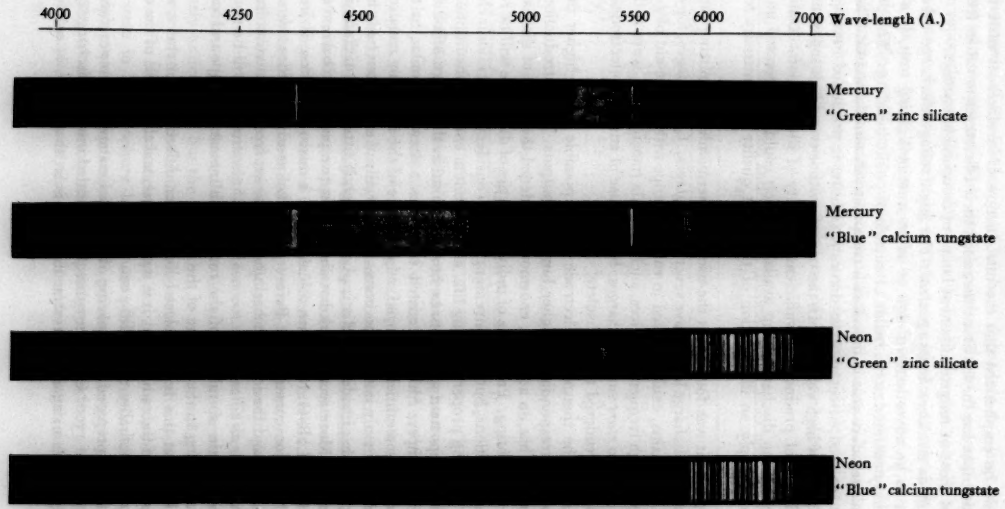


Figure 5. Combined spectra of cold cathode discharge tubes and luminescent powders.

It should be emphasized that the introduction of luminescent materials into cold cathode tubes has for the first time raised their efficiencies to values frequently in excess of those of tungsten filament lamps of comparable wattage.

Hot cathode discharge lamps: general. The high voltage discharge tube, when used either as an advertisement sign or as an interior lighting unit, is a feature almost entirely confined to the inside or outside of public buildings. The provision of the requisite safety gear is therefore a comparatively simple matter. The use of such tubes for road, street, or even domestic lighting is precluded because of the inherent danger of high voltages. In the case of road or street lighting, there is the added difficulty of producing a sufficient intensity of illumination. Until about 1930 there was no discharge lamp available which could be operated on mains' voltages and supply an illumination of the right quality and intensity for these purposes.

During 1930 it was found that the use of a thermionically emitting cathode in various types of discharge lamps gave several advantages. Low voltages, for example the 200 V. a.c. mains, could be used to run the lamps without the aid of a step-up transformer. High frequency devices, such as Tesla transformers, were at this time frequently used to *start* such discharges, but the use of auxiliary electrode systems has now largely rendered these obsolete.

The use of the thermionic electrode makes possible much higher discharge currents with correspondingly higher luminous outputs. The introduction of the low voltage tube has, in addition, enormously widened the scope of discharges for many types of lighting. Properties of individual types of discharge will be reviewed in subsequent sections, but it is convenient here to consider the thermionic electrode.

Steady running properties, long life, and minimum energy consumption in the tube are three important aspects to be borne in mind in the design of an electrode for discharge lamps. Although coated electrodes somewhat similar to those of radio-valves are sometimes employed, an electrode of the following construction is rather more commonly in use. This consists essentially of a tungsten helix mounted on leading-in wires; inside the helix is placed a tightly fitting rod of thermionically active material. More recent work has shown that two particular barium silicates, $3BaO.1SiO_2$ and $2BaO.1SiO_2$, are valuable, and a mixture of these is frequently used as the major constituent. Before this, alkaline-earth oxides were almost invariably used, and the idea that barium could be set free for emission from such stable compounds as silicates was a considerable innovation. It will be understood that the size of the spiral and of the rod of alkaline-earth compound vary considerably according to the wattage of the lamp.

It is clear that this type of electrode is rather different from the more usual coated filament of the radio-valve; it is more robust, and the details of the emission mechanism are probably rather different.

The rods are prepared by furnacing of extruded material to temperatures in the neighbourhood of 1500° C. When subsequently sealed into the discharge tube, the electrodes receive a degassing treatment which is in the nature of an activation process.

The progress of the last few years has been dependent to a large extent on the discovery of new glasses. This applies particularly to the metal vapour lamp. Sodium vapour lamps, with envelopes of soda-lime-silica glass, usually last no more than an hour or so, the failure being due to extreme blackening caused by reductions of silica. It was found that this could be avoided by the use of a special glass containing only a small amount of silica, but a rather large proportion of boric oxide. Many variants on this have since been produced, and it is usual for sodium-resistant glasses to contain anything up to 20 per cent of alumina. Table 2 below summarizes the compositions and some of the properties of a number of glasses used in sodium vapour lamps[2].

Table 2. Composition and physical properties of glasses used for sodium vapour discharge lamps

Glass no.	S 1	S 2	S 3	S 4	S 5	S 5	S 6	S 7
SiO_2	25	21·5	25	22	60·5	10	25	—
Al_2O_3	10	22	30	20	7	27	15	27
CaO	11	10	10	10	5	10	10	—
B_2O_3	50	40	31	34	11	41·5	50	39
MgO	6	—	—	10	3·5	—	—	—
Na_2O	—	6·5	4	4	6	6	—	8
K_2O	—	—	—	—	5	0·5	—	—
BaO	—	—	—	—	2	5	—	26
Coefficients of thermal expansion × 10^6								
	5·8	6·3	6·0	6·1	7·2	6·4	4·9	*ca.* 8·5
Softening point (viscosity = 10^8 poises)								
	—	650°	720°	710°	700°	680°	700°	600°

The first successful sodium lamps were made throughout of the special glass; present lamps are frequently constructed of what is known as "flashed" glass. An ordinary glass and a sodium-resistant glass are drawn into a single tube simultaneously so that the resistant glass forms a thin skin about 0·1 mm. thick on the inside wall; the two glasses have, of course, to be matched in thermal expansion. Another method consists in applying a glaze mixture to the tubing, which is then heated to the softening-point of the glaze. Whatever the method, the resistant glass has to withstand attack from sodium vapour up to temperatures of 300° C.

In addition to the properties already mentioned, such glasses as those of table 2 must have an extremely low iron content. In spite of these precautions, the glass becomes slightly coloured during the life of the lamp; the reason for this is not completely understood, but the mechanism is probably similar to that associated with the coloration of alkali halide crystals heated in alkali vapour.

Glasses for low pressure mercury lamps may be of the soda-lime-silica type, but glasses used in high pressure lamps must possess entirely different properties. Extra-high-pressure lamps require quartz envelopes. Lamps running at an atmosphere or so require glasses with the following properties:

(i) The softening point, defined as in table 2, must lie in the neighbourhood of 900° C.

(ii) The specific electrical resistance must be at least 2 megohms per cm³ at 600° C.

(iii) The iron content of the glass must be low, because the glass must be reasonably colourless without the use of decolourizing agents.

Many variations in composition are possible, but the properties required can be obtained when the silica content is about 50 per cent; when the calcium oxide plus zinc oxide proportion is at least 16 per cent; when the boric oxide content is less than 10 per cent and the alumina content is about 20 per cent.

Table 3 below gives more detailed information with regard to 10 glasses suitable for mercury discharge lamps running at temperatures of several hundred degrees.

Table 3. Composition and physical properties of some glasses used for mercury vapour discharge lamps

Glass no.	1	2	3	4	5	6	7	8	9	10
SiO_2	50	53	40·9	29·3	39·4	51	45·4	54·5	44·7	52·5
Al_2O_3	20	20	19·7	21·1	24·2	12	19·8	21·0	20·7	23·2
CaO	14	15·4	21·3	25·4	18·2	15	13·1	15·2	14·8	8·3
MgO	8·25	—	—	5·2	—	9·25	2·9	—	—	4·2
B_2O_3	5	8·2	10·1	1·8	10·7	10	12·3	7·5	10·6	4·1
BaO	0·5	3·2	—	4·9	—	0·5	4·2	0·8	—	—
Na_2O	1·5	—	—	—	—	0·75	—	1·3	—	—
K_2O	0·75	—	—	—	—	1·5	—	—	—	—
ZnO	—	—	8·3	5·2	8·3	—	0·8	—	9·3	8·2
P_2O_5	—	—	—	7·1	—	—	—	—	—	—
Softening point	905°	920°	815°	810°	825°	830°	760°	885°	825°	930°
Linear coefficient of thermal expansion $\times 10^6$	—	4·7	6·2	7·3	5·1	5·6	4·9	5·1	5·15	4·2
$\log_{10}$ specific electrical resistance at:										
300°	11·1	12·5	12·3	12·0	12·4	11·3	11·9	10·7	12·9	11·6
400°	9·6	11·0	11·1	10·6	11·1	9·8	11·1	9·0	11·1	10·4
500°	8·4	9·6	9·8	9·4	9·9	8·5	9·7	7·8	9·7	9·0
600°	7·4	8·6	8·7	8·2	8·6	7·6	8·5	6·9	8·5	8·0
700°	6·6	7·8	7·5	7·3	7·5	6·5	7·3	6·2	7·3	7·2

Metallic vapour lamps for laboratory purposes. There are many investigations for which small, convenient, and constant line spectrum sources are essential: problems requiring such sources frequently occur in the fields of polarimetry and interferometry. As a result of the work which has just been described, many new and useful lamps are now obtainable; and the earlier and inadequate methods of obtaining as constant sources the arc spectra of the elements zinc, cadmium, mercury, sodium, potassium, rubidium, caesium and thallium need no longer be used.

The general construction of this class of hot cathode lamp is indicated both in the photograph of figure 6 and the circuit diagram of figure 7. The discharge tube is surrounded by a jacket, usually evacuated, through which the four connexions of the electrodes are led out to an ordinary radio-valve cap. Each lamp in addition to a supply of the metal in question contains a rare gas at low pressure to enable the

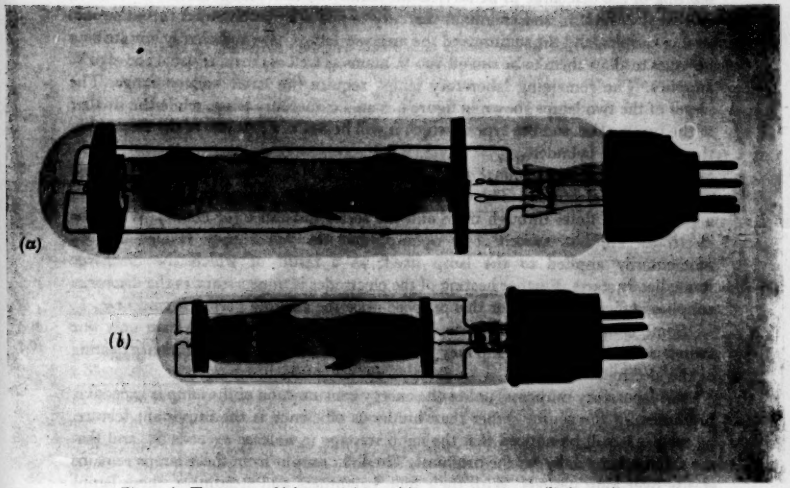


Figure 6. Two types of laboratory lamp: (*a*) a.c. mercury type; (*b*) d.c. sodium type.

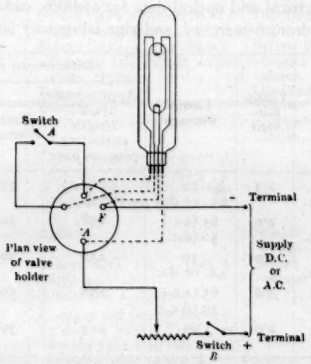


Figure 7. Circuit suitable for laboratory discharge lamp.

starting of the discharge to be carried out readily on low voltage circuits. Sodium, mercury, cadmium, and cadmium-mercury lamps may be obtained for either a.c. or d.c. circuits, and the sodium and the mercury lamps have sufficiently low striking voltages to allow them to be run off 100 V. mains as well as the more usual 200–250 V. supplies. The remaining laboratory lamps require the latter voltage range. The larger of the two lamps shown in figure 6 is an a.c. mercury lamp, while the smaller of the two is a d.c. sodium type in which it will be noticed that one of the electrodes, the anode, is a cylinder.

The circuit shown in figure 7 indicates the method of starting lamps not supplied with auxiliary electrodes. The electrodes are heated by closing switches *A* and *B*; a current then flows through them and the series resistance (or choke). When the electrodes are hot, switch *A* is opened, and the full supply voltage thereby momentarily applied to the lamp itself, as a result of which the discharge immediately starts. Further heating of the electrodes is unnecessary as the discharge supplies sufficient energy for this purpose once the starting period is complete.

Some lamps are supplied with an auxiliary electrode, in which case only one switch is necessary, as the lamp may then be started without any preliminary heating of the electrodes.

For laboratory purposes, unless the energy consumption of the lamp is excessive, brightness of the source rather than luminous efficiency is the important feature. In table 4 it will be noticed that the lamp wattage in no case exceeds 65, and that the sodium lamp is by far the brightest. The light output from these lamps remains steady over long periods. With their high brightness and steadiness, these lamps possess two obvious and outstanding advantages over the earlier flame sources.

Table 4. Electrical and optical data for sodium, cadmium, mercury, cadmium-mercury, and zinc laboratory lamps

Type of lamp	Supply voltage	Lamp current	Lamp wattage	Dimensions in mm.			Brightness of light-source (candles/cm.)
				Approximate overall length excluding pins on base	Overall diameter	Area of light-source	
Sodium	100–250 a.c. or d.c.	1·3	20 a.c. or d.c.	153 (a.c.) 168 (d.c.)	35	16 × 24	7·0
Cadmium	200–250 a.c. or d.c.	2·0	65 (a.c.) 50 (d.c.)	238	50	60 × 20	0·7–1·0
Mercury	100–250 a.c. or d.c.	2·0	30 a.c. or d.c.	238	50	60 × 20	2·0
Cadmium-mercury	200–250 a.c. or d.c.	2·0	65 (a.c.) 50 (d.c.)	238	50	60 × 20	0·7–1·0
Zinc	220–250 a.c.	2·0	—	212	30	25 × 10	—

It remains to say a little about the spectra of some of the laboratory lamps and the filters which may be used to isolate some of the more important lines. In the case of the sodium lamp, the emission of light in the visible spectrum is almost

completely confined to the D lines. This is clearly shown in table 5, which gives values of the relative energies and visual intensities of the lines from 5683 to 6161 A.

Table 5. Relative energies and visual intensities of sodium lines

Wave-length (A.)	Relative energy	Visual intensity
5683 } 5688 }	58	55·8
5890 } 5896 }	4880	3738
6154 } 6161 }	14·7	6·4

Complete isolation of the D lines can be obtained by using the filters suggested at the head of table 6.

Table 6. The isolation of spectrum lines of laboratory discharge lamps

Type of lamp	Wave-length (A.)	Filters	
		For the isolation of lines of column (2)	Additional filter for the suppression of infra-red
Sodium	5890/96	Schott OG 2, 2 mm. thick.	Schott BG 17, 6 mm. thick + copper sulphate cell 10 mm. thick; concentration 57 g./l.
Mercury	3126/32	Schott UG 2, 2 mm. thick + liquid filter of potassium chromate 20 mm. thick; concentration 150 mg./l.	Copper sulphate solution 10 mm. thick; 57 g./l.
	3341	Schott UG 2 + GG 2 each 2 mm. thick + 20 mm. nitric acid $N/5$.	—
	3650/63	Schott UG 2 + Schott BG 12 each 2 mm. thick.	Distilled water, 20 mm. thick.
	4047	Schott GG 4, 1 mm. thick + Corning red purple ultra 2 mm. thick.	Schott BG 17, 6 mm. thick + copper sulphate solution as above.
	4358	Zeiss C.	Distilled water as above.
	5461	Zeiss B.	Schott BG 17, 6 mm. + copper sulphate solution as above.
Cadmium	3261	UG 2, 2 mm. + (solution of 303 g. $NiSO_4$ + 86·5 g. $CoSO_4$ in 1 l. H_2O), 20 mm. thick + K_2CrO_4 filter as for Hg 3126.	—
	4678/800	Agfa 43.	Schott BG 17, 6 mm. + copper sulphate as above.
	5086	Schott GG 11, 2 mm. + Agfa 44.	Schott BG 17, 6 mm. + copper sulphate as above.
	6438	Schott RG 1, 2 mm.	BG 17, 6 mm.
Zinc	3076	Corning red purple corex + 5 mm. + $NiSO_4$. $CoSO_4$ as for Hg 3261 + 20 mm. filter of picric acid 16 mg./l.	—
	4680 } 4722 } 4811 }	Schott GG 8, 1 mm. + BG 12, 2 mm.	Distilled water, 20 mm.
	6362	Schott RG 1, 2 mm.	BG 17, 6 mm.

It should be noted that the cadmium-mercury lamp provides a series of well-spaced lines throughout the spectrum; such a series is frequently very useful for calibration purposes.

The remaining part of table 6 gives data concerning the filters required to isolate a number of the more important lines of the mercury, cadmium and zinc spectra. The origin of the glass filters is indicated by the name of the manufacturer.

Other types of low pressure lamp. (*a*) *Hot cathode neon arcs.* Hot cathode neon arcs with long positive columns have been in use for floodlighting purposes for about four years[3]; more recently such tubes have also been applied, for example,

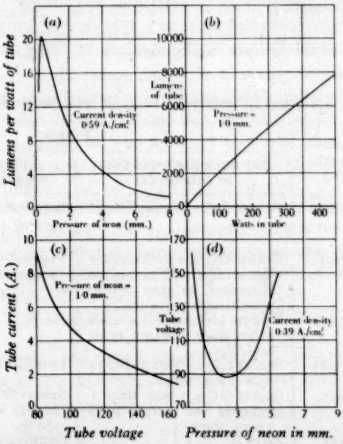


Figure 8. Some electrical and photometric characteristics of hot cathode a.c. neon arcs. The internal diameter of the glass tube is in all cases 3·6 cm., and the electrodes consist of barium oxide rods in tungsten spirals.

to the promotion of more rapid growth and better cropping of tomato plants. Neon tubes, with positive columns from 50 to 100 cm. long and diameters from 2 to 4 cm., provide the most efficient means of obtaining red light-sources. The longer the tube the more efficient it is, since the electrode losses are then a smaller fraction of the total wattage; the lengths mentioned, however, represent roughly the extremes now in use.

The luminous efficiency of a hot cathode neon arc depends chiefly on the length and diameter of the tubing, the nature of the electrodes, the pressure of the gas, and the current density. The efficiency of the *positive column* is, of course, independent of the length of the tubing, except for very short lengths and low pressures; to refer to the efficiency of the column is frequently misleading. Figure 8 shows the effect of (*a*) pressure on tube efficiency; (*b*) the relation between luminous

output and tube watts; (*c*) the current voltage characteristic; (*d*) the voltage-pressure curve. It must be emphasized that a change in either current density, pressure, or diameter will produce a new curve in each case. The accuracy of the photometric measurements on which these curves are based is ± 10 per cent; the method of measurement has been described in an article by Winch and collaborators [4].

It is not easy to make neon arcs which will run off the ordinary mains supplies at pressures much below 2 mm. Most commercial tubes are therefore filled to approximately 2·0 mm. pressure. As distinct from the experimental data of figure 8, the following figures refer to two of the tubes obtainable commercially.

Table 7. Efficiencies of two commercial hot cathode neon tubes

Length of positive column (mm.)	Diameter (mm.)	Tube wattage	Overall wattage	Tube current	Estimated total output (lumens)	Estimated tube efficiency (lumens/watt)	Estimated overall efficiency (lumens/watt)
430	36	150	175	2·8	1500	10·0	8·5
865	36	400	470	6·0	4450	11·0	9·5

Neon tubes require, in addition to the series choke, a filament transformer and a starting device for successful operation. The starting device may take the form of a high frequency coil operated from an additional winding on the cathode heating transformer; or it may be an auxiliary electrode system within the tube itself. The 150 watt lamp of table 6 is started by the latter method.

(*b*) *Low pressure mercury hot cathode tubes.* Low pressure mercury lamps, very much larger than the laboratory lamp already discussed, are now a common feature of coloured floodlighting installations, and are generally similar in size to the neon lamps. On account of the low pressure of mercury vapour at room temperatures, argon is usually employed as a starting gas. Except in so far as it affects the mercury

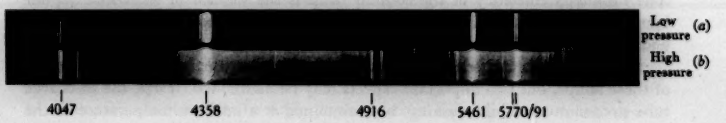


Figure 9. Spectra of mercury discharge lamps.

vapour pressure, the argon probably has no bearing on the luminous efficiency of the discharge. The spectrum of a clear glass lamp of this type is shown in figure 9 (*a*); it is essentially the same as that of the cold cathode mercury tube in clear glass. Other colours are obtained by the use of different glasses; for example, a dark blue light is produced by the use of blue glass, a green by the use of "uranium" glass. In the latter case the glass is fluorescent, so that there is emission as well as selective absorption.

Table 8 includes efficiency data for four colours, blue, pale green, green, and dark blue, obtainable commercially.

Table 8

Colour	Diameter (mm.)	Length of positive column (mm.)	Tube wattage	Overall wattage	Estimated total output (lumens)	Estimated tube efficiency (lumens/watt)	Estimated overall efficiency (lumens/watt)
Light blue	26	430	100	130	1250	12·5	9·5
Dark blue					550	5·5	4·25
Light green					1250	12·5	9·5
Dark green					950	9·5	7·25
Light blue	26	865	250	300	3600	14·5	12·0
Dark blue					1650	6·6	5·5
Light green					3600	14·5	12·0
Dark green					2700	10·75	9·0

Photometric investigation has shown that the shortest wave-length to be transmitted by the clear glass is the line 3132 A.; there is thus no harmful short wave-length radiation emitted by these lamps. In the case of the green lamp no ultra-violet lines are transmitted and the blue and violet lines are almost completely absorbed. For equal wattages it has been shown that the ratio of the total energies of the neon and mercury lamps emitted in the visible spectrum is about 1·5; the ratio of the total visual intensities is approximately 0·8.

(*c*) *The sodium vapour lamp*. The following remarks apply to the sodium lamps used for road and floodlighting purposes, and, as might be expected, these have considerably higher wattages and luminous outputs than the laboratory type. With these also, the vapour pressure is low and in the neighbourhood of 0·007 mm. Whereas brightness was important for the laboratory lamp, high efficiency and luminous output are the chief objective when the lamps are required for road or floodlighting. The sodium vapour lamp is also a low pressure type of discharge. The luminous efficiency of the sodium lamp is the highest of the metal vapour lamps in common use. The wattages of the lamps available range from 50 to 150, and average tube efficiencies from about 43 to 65 lumens per watt.

It is impossible to discuss in detail the many factors which influence the design of these lamps, but two important aspects may be mentioned. First, the discharge tube proper must, while running, be maintained at a uniform temperature in the neighbourhood of 300° C. This is usually accomplished by placing the lamp inside a Dewar flask; sometimes the space between the lamp and the Dewar is evacuated. Secondly, the effects of current-density and of self-absorption of the resonance lines determine to a large degree the diameter of tubing employed for a given wattage lamp. Figure 10 indicates diagrammatically the general construction of a 100 watt lamp used for floodlighting; in another design the discharge lamp is in the form of a U-tube.

The low pressure of sodium vapour at room temperature necessitates the addition of a rare gas for starting, and neon is usually employed. Starting of the

linear lamp on mains' voltages is accomplished by means of an internal auxiliary electrode system; the U-tubes require higher voltages, and strayfield transformers are used. The brightness values of these lamps lie between 10 and 20 candles/cm^2. Articles by Fonda and Young[5], Found[6], and Heller[7] will be valuable to those further interested in sodium lamps.

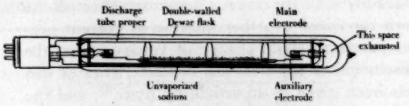


Figure 10. 100 watt sodium floodlighting lamp (length of discharge about 22 cm.).

High pressure mercury vapour lamps. At the present time there are several high efficiency mercury sources in use or in development, and it is convenient to classify them according to the brightness of the source when the discharge is in a steady running condition. The low pressure tubes we have already discussed, and in these the brightness is about 2·0 candles/cm^2.

Table 9. Brightness values of high pressure mercury lamps

Type	Brightness (candles/cm^2)
400 W. high pressure mercury (glass)	170
250 W. high pressure mercury (glass)	160
150 W. high pressure mercury (glass)	150
125 W. and 80 W. high pressure mercury (quartz)	750–1000

The remaining lamps fall into the following classes. First there are the street lighting types, which were among the pioneer lamps of the hot cathode class, and the brightness of these is in the neighbourhood of 150 candles/cm^2. For smaller wattages, it is necessary to use quartz, and the brightness figures are then 750–1000 candles/cm^2. Finally, a number of special high brightness lamps are under development. The brightness of these sources varies inversely as the life of the lamp, and although it has been possible experimentally to obtain brightness values of 180,000–200,000 candles/cm^2, such figures[7] can only be obtained for short intervals. Lamps with brightness values of 25,000 give longer lives.

The main features of the high pressure street lighting lamps (glass) are indicated in figure 11; the lamp consists essentially of a tubular glass vessel *d* with electrodes *e*, *e*, of the kind described earlier in this article, sealed into each end. After a suitable evacuation of the tube and activation of the electrodes a few mm. of argon and a small measured quantity of mercury are admitted to the lamp, which is then sealed off. The lamp is then mounted on the pinch *p* and the auxiliary electrode *a* of tungsten connected through a high resistance in the well of the pinch to the main electrode at the opposite end of the lamp; the purpose of *a* is simply to facilitate starting on supply voltages of 250 or less. The mounted lamp is then sealed into the outer jacket *j*. When the discharge is first started, it is of the low pressure

type; as the lamp warms up, so the mercury vapour pressure increases, and the discharge finally takes the form of a constricted column co-axial with the lamp when it is in a vertical position. The quantity of mercury is restricted so that complete vaporization is achieved when the discharge is in the steady state. During the starting period the voltage across the lamp rises steadily with the vapour pressure; when all the mercury has been vaporized a further increase of current produces only a comparatively slight change of voltage across the lamp. A full discussion of the electrical characteristics of this type of lamp has been given in an article by Ryde[8], and the reader is referred to this for further details on this aspect of the subject. It will be realized that cumulative ionization renders the use of stabilizing impedance absolutely necessary with these, as with other, discharge lamps, and the design of these has been one of the major problems connected with the successful application of the lamps. The power factor of a discharge lamp proper may lie between 0·7 and 0·95, and the value for the whole circuit when the choke is in series is frequently very much below 0·5. In order to correct this, condensers are placed across the mains; in the case of the 400 W. street lighting lamp 20 μF. are required; the 400 W. neon lamp, referred to in the low pressure section, requires 80 μF. to bring the power factor to a value of 0·85.

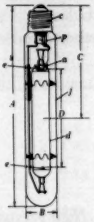


Figure 11. The high pressure mercury lamp in glass.

This high pressure mercury lamp originated in England in its present practical form, but it should be remembered that the effect of pressure on efficiency had been observed in a very elementary way as early as 1906[1]. The first installation of the new lamps was erected along a length of test road at Wembley in 1932. High pressure mercury vapour lamps, essentially the same as those developed in England, are now extensively used on the continent and in America.

Table 10

Lamp wattage	Lumens per watt at 0 hours	Average lumens/watt over 1500 hours
150	32	26
250	36	31
400	45	37

Three wattages are now in common use, and the efficiencies of these are shown in table 9; during life, blackening gradually occurs at the ends of the discharge tube proper, and the average efficiencies over a period of 1500 hours are shown in the third column. 1500 hours is usually regarded as the economic life.

When viewed through a dark glass the column of light appears much narrower than the tube; this is a genuine pressure effect, but careful measurements show that the luminous column is a good deal wider than a casual glance would lead one to suppose. Figure 12 illustrates this point by showing the distribution of brightness across the arc.

A little must now be said about the spectra of these lamps. The spectrum of a high pressure lamp differs a good deal from that of the low pressure variety. Figure 9 (*a*) and (*b*) compare the spectra of the two types. It will be noticed that, although the violet lines are not very different in intensity, the green and yellow lines are very much stronger in the high pressure tube, which has, in addition, a background of continuous radiation and a number of fainter lines in the green, yellow and red.

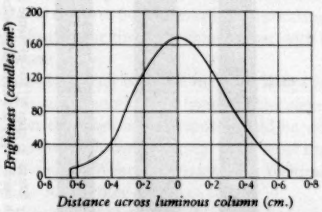


Figure 12. The variation of brightness across the luminous column of a high pressure 400 watt mercury vapour lamp (in glass).

Energy values for a large number of lines in the high pressure mercury spectrum have been determined by Krefft and Pirani[9] for a number of different pressures. Their measurements were made on a discharge in a quartz envelope, as a result of which a number of lines not normally transmitted by the glasses used for the lamps of table 10 were observed. The energy values of table 11 have been interpolated from Krefft and Pirani's results, and they are relative to the composite figure of 100 for the triplet 3650/63. Lines 2358–3026 are completely absorbed by the glass lamps of table 10; lines 3126–3341 appear faintly when the discharge tube is not surrounded by an outer jacket. Line 10140 is, of course, in the infra-red. The relative energies of the visible spectrum must be multiplied by the appropriate ordinate of the luminosity curve in order to transform them into relative visual intensities.

Table 11. Relative energies in the spectrum of the high pressure mercury lamp at a pressure of 760 mm. (after Krefft and Pirani)

Wave-length (A.)	Relative energies λ 3650/63 = 100	Wave-length (A.)	Relative energies λ 3650/63 = 100	Wave-length (A.)	Relative energies λ 3650/63 = 100
2358	3·0	2699	5·3	3906	1·5
2378	5·3	2753	3·9	4047	27·0
2400	5·0	2804	12·5	4078	5·4
2464	2·1	2894	6·1	4358	50·1
2483	12·3	2925	2·5	4916	1·3
2537	17·0	2967	16·0	5461	64·6
2576	5·6	3022/26	32·8	5770/91	78·7
2603	2·7	3126/32	63·5	6907	2·0
2640	3·3	3341	8·6	10140	36·5
2652	24·2	3650/63	100·0		

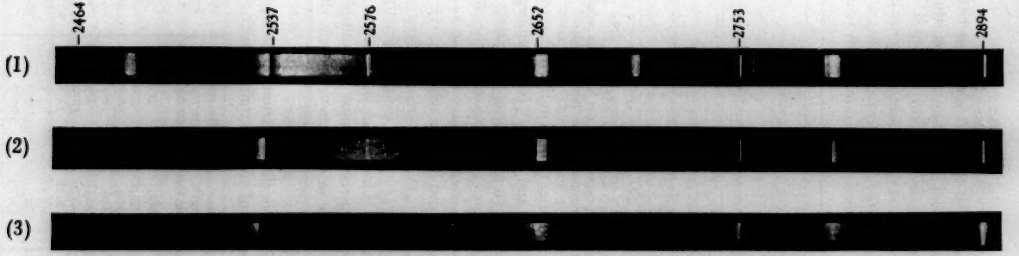


Figure 13. Illustrating the effect of pressure on the U.V. spectrum of high pressure mercury lamps in quartz; the pressure increases from (1) to (3). (This part of the spectrum is not transmitted by the glass bulbs in which such lamps are normally enclosed.)

It is common knowledge that the light from high or low pressure mercury lamps is deficient in red radiation, and much effort is being devoted to the improvement of colour in this respect. These efforts, have, broadly speaking, taken two courses; modification of the spectrum by the addition of cadmium to the mercury, and increase of red radiation by means of luminescent powders on the inside of the outer bulb. By using mercury-cadmium high pressure lamps in conjunction with suitable luminescent powder, the percentage red can be still further increased to a value in the neighbourhood of 5·0 to 6·0 *. Lamps incorporating cadmium have been available for some time, but, at the time of writing, those with luminescent powders are still in the development stage.

The outstanding feature of high pressure mercury lamps is their high luminous efficiency compared with tungsten filament lamps of the same wattage. It is most important, therefore, that this high light output should be utilized to the utmost; so much may be lost in a badly designed fitting. This most important aspect of the problem has been discussed by Wilson, Damant, and Waldram[10], and the reader is referred to their article for further details.

A portion of the u.v. spectrum of three different higher pressuré lamps in quartz is shown in figure 13, and two interesting features may be pointed out. First, there is a good deal more continuous radiation in the spectra than there is in those of figure 9, and some of this lies in the infra-red. Second, as the pressure increases from (1) to (3), an increasingly wide absorption band appears in the neighbourhood of line 2537 with an extremely sharp cut-off on the short wave-length side.

* The "percentage red" is defined as the percentage transmitted by Wratten Filter No. 25.

References

(1) Ryde, J. W. *J. Roy. Soc. Arts*, **82**, 624 (1933).
(2) Partridge, J. H. *J. Soc. Glass Tech.* **19**, 266 (1935).
(3) Ackerley, R. O. *G.E.C. Jl.* **5**, No. 4 (Nov. 1934).
(4) Winch, G. T., Palmer, E. H., Machin, C. F. and Dudding, B. P. *G.E.C. Jl.* **5**, No. 3 (1934).
(5) Fonda, G. and Young, J. *J. opt. Soc. Amer.* **24**, 31 (1934).
(6) Found, C. G. *Gen. elect. Rev.* (June 1934).
(7) Heller, G. *Philips Tech. Rev.* **1**, 2 (1936).
(8) Ryde, J. W. *G.E.C. Jl.* **4**, 199 (1933).
(9) Krefft, H. and Pirani, M. *Z. tech. Phys.* **14**, 393 (1933).
(10) Wilson, G. H., Damant, E. L. and Waldram, J. M. *J. Instn elect. Engrs.*, N.Y. **79**, 241 (1936).

X RAYS

By THE X-RAY STAFF, PHYSICS DEPARTMENT, NATIONAL PHYSICAL LABORATORY

General Editor: G. W. C. Kaye, O.B.E., M.A., D.Sc.

§ 1. PRODUCTION OF X RAYS

By G. E. Bell, M.Sc.

Until comparatively recently, high-tension generators and X-ray tubes were not made for operating at voltages exceeding about 250 kV. During the last few years, however, great advances have been made, and X rays excited at upwards of 1000 kV. have been generated and used for both medical and physical purposes. In the earlier developments, sealed-off X-ray tubes for 300, 400 and later 600 kV. were made on the lines of the conventional self-protected tubes. These tubes, which were of the type having both electrodes insulated, proved very successful, and a number of hospitals and research institutions now use 400-kV. X rays for routine work. The necessary high-tension generators are readily available, usually in the form of a number of generators giving a lower potential connected in series.

Various attempts to make tubes operating at still higher voltages have now come to fruition, thanks largely to improvements in vacuum pumps. Lauritsen and Bennet[1] made an experimental 750-kV. continuously-evacuated tube, and later Lauritsen and Casson[2] improved this tube to such an extent that regular work became possible. In operating X-ray tubes at very high voltages a number of difficulties are encountered, chief among them being the prevention of cold emission of electrons (i.e. emission due solely to intense electric fields), the proper distribution of the potential gradient in the tube and the obtaining of the necessary high vacuum. Cold emission is prevented by making the electrodes of sufficient size and giving them suitable curvatures, while the potential gradient along the tube is usually correctly balanced by designing the tube and high-tension generator so that various parts of the tube can be connected to points on the generator at suitable potentials. The vacuum necessary for successful operation depends largely on the design of tube. If the anode-cathode spacing is short, a comparatively low vacuum (10^{-4} mm. mercury) is sufficient, while if the tube has a long electron path, a feature which reduces trouble due to cold emission, a much better vacuum (10^{-6} mm. mercury for an electron path of 10 ft.) is necessary. In any case the overall length of the tube must be great—of the order of 1 ft. per 100 kV. potential—and the diameter large (not less than 1 ft.) to reduce the electrical stresses in the glass or porcelain envelope.

The high-tension generator may give an output voltage which is either constant or pulsating. In the latter case, the filament of the X-ray tube is set in a cylindrical

recess in the cathode, where the field-strength is so low that electrons only flow at or near the peak voltage of the supply.

Lauritsen's original tube had an earthed anode, contained in a metal lead-covered discharge-chamber at earth potential. This arrangement, although offering several features of practical convenience, is unsuitable for voltages exceeding about 750 kV. peak, and later tubes described by Lauritsen and his collaborators were of

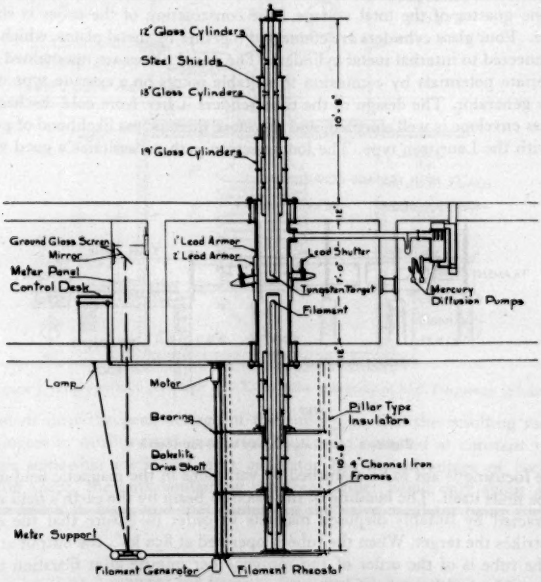


Figure 1. Lauritsen X-ray tube for 1000 kV.

the double-ended type having a central earthed discharge-chamber typified in the diagram of figure 1. Difficulties of cooling the anode were overcome by using forced water cooling, the water passing through a motor-car radiator, contained in a suitable corona shield. A tube of this type, installed at the Kellogg Radiation Laboratory, Pasadena, has been used for clinical purposes since 1932, operating at 900 kV. and 3 milliamperes for many hours continuously. The filament requires replacement every few weeks, but after this operation the tube can be evacuated and put into service again in about 6 hours. The X-ray output of the tube, measured at 57 cm. from the tube and after filtration through 1 mm. of lead and 5 mm. of steel

was about 15 rontgens per minute. X-ray tubes of this general design have been installed in a number of American hospitals and research institutions, and it was with such an equipment that Read investigated the absorption of X rays having wave-lengths of 20–50 X.

Tubes of a different pattern have been designed by Coolidge[3] and installed at the Swedish Hospital, Seattle, and at the Mercy Hospital, Chicago. The distinctive feature of these tubes is that the electrons are accelerated in four stages, each one quarter of the total voltage. The construction of the tubes is shown in figure 2. Four glass cylinders are connected together by metal plates, which in turn are connected to internal metal cylinders. The metal plates are maintained at their appropriate potentials by connexion to suitable points on a cascade type of high-voltage generator. The design of the tube renders it free from cold discharge, and the glass envelope is well shielded, and therefore there is less likelihood of puncture than with the Lauritsen type. The long electron path necessitates a good vacuum,

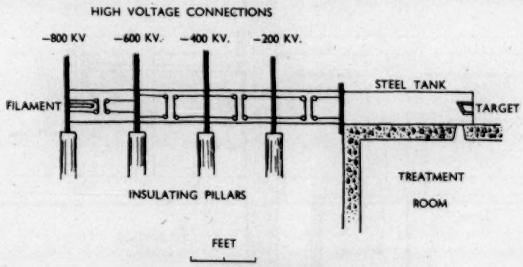


Figure 2. Coolidge X-ray tube for 1000 kV.

and the focusing is apt to be disturbed by variations in the magnetic field in which the tube finds itself. The bending of the electron beam by the earth's field must be counteracted by suitably disposed magnets in order to ensure that the electron beam strikes the target. When the tube is operated at 850 kV., the output at 50 cm. from the tube is of the order of 120 rontgens per minute after filtration through 2·7 mm. of lead and 3 mm. of brass.

A high-tension generator and X-ray tube of very unorthodox design operating at 900–1200 kV. has been constructed by Sloan[4] at the University of California. Essentially a development of the Tesla transformer, the high-tension generator (figure 3) comprises a valve oscillator (operating at a frequency of about $5 \cdot 10^6$ c./sec.) to which is coupled a secondary winding made of concentric copper tubes and tuned to synchronism. The secondary winding carries at its high-tension end a tungsten disc which acts as the target of the X-ray tube and is cooled by water flowing through the copper tubing. Both the primary and secondary windings of the Tesla transformer are enclosed in a lead-shielded metal tank which is evacuated by means of oil-diffusion pumps. A hot filament contained in a tubular holder in the side of

the tank serves to provide the necessary electrons which, accelerated by the high-frequency high-voltage field, impinge on the target and produce X rays. A biasing voltage of some 10,000 V. is applied to the filament to ensure that current only

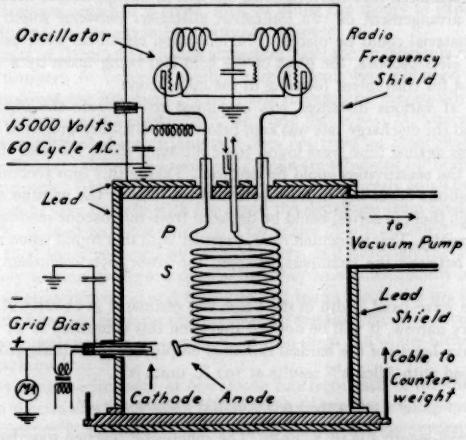


Figure 3. High-tension generator and X-ray tube operating at high frequency (Sloan)

passes at or near the peak voltage of the supply, so that the resulting radiation approximates in quality to that emitted by a tube operated at constant voltage. Although somewhat inefficient, this generator has the advantage of occupying very little space, the tank containing the transformer being only about 4 ft. in diameter and 4 ft. high, and that containing the valve oscillator being somewhat less.

References

(1) Lauritsen and Bennet. *Phys. Rev.* **32**, 850 (1928).
(2) Lauritsen and Casson. *Phys. Rev.* **36**, 988 (1930).
(3) Coolidge. *Amer. J. Roentgenol.* **27**, 405 (1932).
(4) Sloan. *Phys. Rev.* **47**, 62 (1935).

§ 2. ABSORPTION OF X RAYS

By I. Backhurst, M.Sc.

The X-ray absorption in lead, tin, copper and iron has recently been measured by Read[1] for the wave-length range 20–50 X. units (1 X. unit = 10^{-11} cm.). Potentials up to 1000 kV. were needed on the X-ray tube to secure adequate radiation intensity at these short wave-lengths. After filtration of the rays by

1 mm. of lead and 10 mm. of steel, a beam of the desired wave-length was selected by means of a spectrometer having a rocksalt crystal. The selected beam passed through a lead diaphragm system inserted in a 10-in. concrete wall; this served to shield from scattered rays the measuring portion of the apparatus, which consisted of a series arrangement of two ionization chambers between which a sheet of absorbing material could be placed. The ionization chambers were used similarly to gold-leaf electroscopes, the place of the gold leaf being taken by a quartz fibre having a foot for convenient focusing in the microscope.

Radium at various distances was employed to calibrate the chambers and, provided that the discharge rate was kept below a certain limit, graphs of microscope scale readings against time were found to be the same shape for each chamber and the ratio of the sensitivities could be deduced. Taking this into account, the ratio of the ionizations occurring in the two chambers, during the passage of an X-ray beam through them in series, could be deduced from microscope readings taken for both concurrently. This ionization ratio, together with that found when an absorber was placed between the chambers, enabled the absorption coefficient to be calculated.

The true wave-band width in the beam was estimated at about 5 X. units but, although very narrow, it will be noticed that even this amounted to 25 per cent of the mean wave-length of the hardest radiation employed. A formula linking up the results for lead with Allen's[2] results at 102 X. units is

$$\tau = 0{\cdot}01377\lambda + 0{\cdot}0003326\lambda^2 + 0{\cdot}00003474\lambda^3,$$

where the wave-length λ is in X. units. The conclusion reached was that using the Klein-Nishina[3] formula for σ (the part due to scattering of the total absorption coefficient μ), and employing the theory given by Hulme, McDougall, Buckingham and Fowler[4] to obtain τ (the part of μ due to true absorption), it appears possible, provided that electron-pair production is inappreciable, to calculate μ for any element for $\lambda < 36$ X. units with an error not greater than 10 per cent.

A wave-mechanical theory of X-ray absorption has been put forward by Blockinzew and Halperin[5], according to which $(\tau/\lambda^5\rho)^{\frac{1}{3}} \equiv M$ should be nearly a linear function of $1/\lambda$, τ/ρ being the true mass absorption coefficient. The linearity does not hold for values of λ near an absorption edge, but a straight line may be drawn cutting the $1/\lambda$ axis where $\lambda = \lambda_K$, the K absorption edge, and passing close to a number of values of $(\mu/\lambda^5\rho)^{\frac{1}{3}} \equiv M'$ calculated from known experimental values of the total mass absorption coefficient μ/ρ in the K absorption region. For sufficiently large values of λ, σ/ρ is small compared with τ/ρ, so that approximately $\tau/\rho = \mu/\rho$, and it is in this region that the theory requires M' to vary linearly with $1/\lambda$. From the extent of departure from linearity at short wave-lengths, σ/ρ is deduced. The experimental values used to check the theory are those of Grösskurth[6] for Al, Ni and Cu and of Allen[2] for Al, the agreement shown being regarded as good confirmation of the theory.

Although it is true that a measure of agreement may be obtained in the instances cited, a proper comparison with experiment cannot really be made, however, until

the theory has been worked out in more detail. If Allen's results for Cu, for example, are plotted with values of M' as ordinates against $1/\lambda$, the graph is a curve concave upwards, and the process of obtaining the best fit with a straight line to a selected portion of the curve is one allowing much latitude, even if the intersection with the $1/\lambda$ axis is regarded as fixed; also, in fitting the theory to experiment, the scale constant has at present to be arbitrarily adjusted.

The results of Grösskurth, which altogether were for 16 elements ranging from carbon to tungsten in order of atomic number and were obtained following a method employed by Küstner[7], have been represented empirically in another manner by Küstner and Grösskurth[8], who find that $\tau/\rho = c\,\lambda^n$, where c and n are constant for a given element as long as λ does not cross an absorption edge. For most of the elements that were examined, c and n are tabulated for both sides of the K absorption edge. In the range of λ within which τ/ρ is nearly equal to μ/ρ, $\log \mu/\rho$ is thus found to vary linearly with $\log \lambda$ and the departure from linearity at short wave-lengths is used to determine σ/ρ. The values of σ/ρ so found, however, are much less than those reported by other workers, and although the authors state, unfortunately without giving any details, that they have subsequently measured the scattered radiation from lead and tantulum and found results for σ/ρ in agreement with those deduced in the above manner, it must remain difficult to accept these new values for σ/ρ until the wide discrepancy between them and values hitherto available has been explained.

By measuring the intensity of one strong line from each of the three L groups of emission spectra of tin and using values of the L absorption-jump ratios for silver due to Kellström[9], Schwarz[10] has deduced for tin the proportion of the total absorption coefficient which may be attributed to each of the L electron groups, when the wave-length of the incident radiation is 2·285, 1·54 or 0·71 A. The effective incident radiation employed was the difference between the radiation obtained when first one and then the other of two "balanced" filters were used in the primary beam from the X-ray tube, the method being that due to Ross[11]. The relative intensities of the emission spectra were measured by means of a Geiger counter. Correction was made for absorption in the specimen. The results showed that the relative effect of the deeper seated electron groups L_1 and L_2 considerably increased as the wave-length of the incident radiation was shortened.

References

(1) Read. *Proc. roy. Soc.* A, **152**, 402 (1935).
(2) Allen. *Phys. Rev.* **27**, 266 (1926).
(3) Klein-Nishina. *Z. Phys.* **52**, 853 (1929).
(4) Hulme, McDougall, Buckingham and Fowler. *Proc. roy. Soc.* A, **149**, 131 (1935).
(5) Blockinzew and Halperin. *Phys. Z. Sowjet*, **7**, 175 (1935).
(6) Grösskurth. *Ann. Phys.*, Lpz., **20**, 97 (1934).
(7) *Reports on Progress in Physics*, p. 307, Physical Society (1935).
(8) Küstner and Grösskurth. *Strahlentherapie*, **52**, 115 (1935).
(9) Kellström. *Z. Phys.* **44**, 269 (1927).
(10) Schwarz. *Ann. Phys.*, Lpz., **23**, 571 (1935).
(11) Ross. *J. opt. Soc. Amer.* **16**, 433 (1928).

§ 3. SCATTERING OF X RAYS

By I. Backhurst, M.Sc.

Measurements of the diffuse scattering of X rays from sodium fluoride, magnesium oxide and silicon carbide have been made by Jauncey and Deming[1] employing the white radiation from a tungsten target tube operated at 34 kV. The intensity of the scattered rays was directly compared with that of the incident beam by limiting the exposure of the latter to 1/60th sec. by means of a pendulum shutter. In order to use the results to test the Waller-Hartree theory[2] of diffuse scattering as applied to neon-like crystals, it was necessary to estimate, for each crystal, the characteristic temperature θ evaluated by Debye[3] in terms of C_v, the specific heat at constant volume. C_v was obtained from known values of C_p, the specific heat at constant pressure, by means of the formulae $C_p - C_v = 0{\cdot}0214\, C_p^2\, T/T_m$ and $C_p - C_v = \beta^2 MT/K\rho$, where T is absolute temperature, T_m is melting point, β is coefficient of volume expansion, M is molecular weight, K is compressibility and ρ is density. Values of θ were found corresponding to data for different values of T, and θ then plotted against θ/T. At low values of T, a limiting value of θ was found and this was taken as the one appropriate to the scattering theory, good agreement with the Waller-Hartree theory being obtained when this was done. These results therefore confirm the deduction made by Harvey, Williams and Jauncey[4], from experimental results due to Williams[5], that the Pauli exclusion principle must be taken into account in X-ray scattering theory.

Discontinuities of a kind not anticipated on current theory have been discovered by Laval[6] in the angular variation of intensity of scattered radiation. Characteristic rays, selected by means of crystal reflection, were allowed to fall on crystalline powders of different metals and the scattered radiation measured, with ionization chamber and electrometer, at angles less than those corresponding to possible regular reflection from the crystalline particles of the metal. A relation expressed by $\nu \sin(\phi/2) = \alpha N + \beta$ was found between the angular position ϕ of a discontinuity, the frequency ν of the incident radiation and the atomic number N of the metal, α and β being constants for a given value of $\nu \sin(\phi/2)$. Values of $\nu \sin(\phi/2)$ corresponding to discontinuities were found for nine metals ranging from aluminium to gold in atomic number and as many as nine discontinuities were found for aluminium. With CuKα-rays incident on the latter metal, the values of $\sin(\phi/2)$ corresponding to discontinuities were 0·085, 0·110, 0·130, 0·178, 0·236 and 0·282.

As $\sin(\phi/2)$ was increased, the scattered intensity was found to increase except at the discontinuities, where drops in intensity were observed, amounting to as much as 40 per cent for aluminium at the $\sin(\phi/2) = 0{\cdot}110$ discontinuity occurring with CuKα-rays. Results obtained with sodium chloride were in accordance with the inference, derived from the above equation involving atomic number, that the phenomenon was purely atomic and unaffected by chemical combination.

References

(1) JAUNCEY and DEMING. *Phys. Rev.* **48**, 577 (1935).
(2) *Reports on Progress in Physics*, p. 308, Physical Society (1935).
(3) DEBYE. *Ann. Phys.*, Lpz., **39**, 789 (1912).
(4) HARVEY, WILLIAMS and JAUNCEY. *Phys. Rev.* **48**, 365 (1934).
(5) WILLIAMS. *Phys. Rev.* **46**, 83 (1934).
(6) LAVAL. *C.R. Acad. Sci.*, Paris, **201**, 889 (1935).

§ 4. IONIZATION IN GASES

By I. BACKHURST, M.Sc.

The technique of the production and focusing of intense positive ion beams has been described in some detail by Tuve, Dahl and Hafstad[1], who have used as ion source a d.-c. arc, operated with external resistance on 110 V. at about 250 W. input. The arc was constricted by a steel capillary and the ion beam obtained from a 1 mm. diameter probe canal, through which the loss of gas from the arc chamber to the high vacuum space was not more than about 10 cm^3 per hour, a matter of importance if the use of deuterium is contemplated. The probe canal voltage of 100–4000 was supplied by a filtered rectifier unit, while a kenetron half-wave unit was used to obtain the 2–15 kV. required across the focusing gap or electrostatic lens between the probe canal and the next metal tube traversed by the ion beam. The gaps between successive metal tubes traversed by the beam formed an electrostatic lens system, to the axis of which the ion source could be adjusted by means of a double eccentric arrangement of cams. A linear relation holds between the voltage on a lens required to produce a given convergence and the initial voltage of the ions. Cascade tubes having up to 23 electrostatic lenses were used with ion currents reaching 20 microamps. and potentials as much as 1200 kV., nearly all the initial ion current being focused on the target.

To investigate the ionization produced in inert gases by positive alkali ions, Varney[2] has adapted a sensitive method used by Lawrence and Edlefsen[3] for detecting photoionization in vapours. A monel metal cylinder, of length and diameter each about 2 cm., is fitted with an axial tungsten filament, a potential of 1 or 2 V. being applied between them so that, when the filament is heated, a space-charge limited electron stream flows from filament to cylinder. An increase in this electron stream may be brought about by admitting a beam of positive ions to the cylinder through a small hole at one end, since part of the space-charge is then neutralized. A high sensitivity results from the slowness of the positive ions in comparison with the electrons and, in these experiments, a positive ion remained about 30,000 times as long as an electron in a given region of the cylinder, each ion being consequently effective in neutralizing the space-charge of 30,000 electrons. A further advantage of the method was that, if the secondary electrons occasioned by the ion beam could produce any effect at all on the electron stream, it would be negligible in comparison with that of the positive ions. Actually two cylinder units

were used, no ions being admitted to one, which was used as a balance to reduce the effect of small fluctuations in voltage supplies, etc.

The ion beam was generated by means of a platinum filament, coated with a Kunsman[4] catalyst, the ions then being accelerated through two slits by electric fields before bombarding the gas, Ne, A, Kr or Xe, in the space-charge ion detector. Increases in the rate of variation of detector current with ion velocity showed that, with the positive ions Li^+, Na^+, K^+, Rb^+, Cs^+, the critical potential V required to cause ionization of the gas varied from about 70 to over 400 V. It was found better to express the results in terms, not of V, but of $Vm_a/(m_a+m_i)$, where m_a and m_i are the masses of the gas atom and alkali ion respectively. The figures then express voltages proportional in each case to the fraction of the ion's kinetic energy becoming available on collision for altering the internal energies of the two colliding particles. These available energies are then seen to be approximately the same for ionization collisions involving atoms and ions having similar electron configurations. For example, Rb^+ is similar to Kr, while also A is similar to K^+, and the available energy in volts for Rb^+ in A is 43 and for K^+ in Kr is 47, the critical potentials being 135 and 69 V. respectively. The results show also that the available energy necessary for an ionization collision is at a minimum value, both for the ion and atom concerned, when the electron configurations of the two are most nearly the same.

The effect of resonance radiation on the conductivity in Ne and He has been investigated by Smith and Duffendack[5], who circulated the gases used alternately through the electric discharge tube and through a purifying system, great care being taken to work with the purest gas possible. An arc supplied by a heated cathode was maintained at each end of the discharge tube and, in the middle of the tube between the two anodes, two parallel plane nickel electrodes 1 cm. in diameter, with 1 cm. wide guard rings, were placed close together shielded on the outer faces by glass. When the applied difference of potential between the guarded electrodes was small, the current-voltage ratio proved independent of voltage and furnished a measure of the conductivity of the gas. With the electrodes perpendicular to the tube axis no measurable current ($< 10^{-10}$ amp.) was observed, but currents of the order of 5×10^{-9} amp. were obtained with the electrodes parallel to the axis.

The inference was made that the agent giving rise to the conductivity must be produced either in, or very close to, the primary light beam from the discharge and that scattered resonance radiation could have negligible effect. Measurements were also made of the increase of conductivity in Ne or He produced by admixtures of Kr or A and of the decrease of conductivity in He produced by an admixture of Ne. The effects observed were considered best explained by assuming the existence in the main gas of metastable atoms which formed ions by collision with atoms of the added gaseous "impurity". In support of this theory it is observed that Ne, the only gas to decrease the conductivity of He, was the only one having an ionization potential higher than that corresponding to the He metastable states.

On the other hand, detailed arguments, based partly on recent experiments at Oxford, have been put forward by Townsend[6] to show that there is no need for a metastable atom theory in order to explain the increased conductivity occasioned in

monatomic gases by the presence of impurities. When account is taken of the statistical distribution of electron velocities in the main gas, the increase in conductivity may be shown to be made possible by electronic collisions with atoms of the added impurity. An important argument in favour of this view is that, on the metastable atom theory, two collisions with the same atom are required to produce ionization and radiation, so that both rate of ionization and intensity of radiation would be expected at low current density to vary as the square of the current, whereas experimentally the radiation intensity is found to be directly proportional to the current and therefore compatible only with the production of ionization by single collisions.

References

(1) Tuve, Dahl and Hafstad. *Phys. Rev.* **48**, 241 (1935).
(2) Varney. *Phys. Rev.* **47**, 483 (1935).
(3) Lawrence and Edlefsen. *Phys. Rev.* **34**, 233 (1929).
(4) Kunsman. *Science*, **62**, 269 (1925).
(5) Smith and Duffendack. *Phys. Rev.* **47**, 473 (1935).
(6) Townsend. *Phil. Mag.* **20**, 242 (1935).

§ 5. THE ENERGY EXPENDED IN IONIZING AIR

By W. Binks, M.Sc.

The most desirable way of measuring the energy flux in a beam of X rays or γ-rays would be to measure the heat developed in a body which completely absorbs the radiation. As the amount of heat produced is very small the thermal method is inconvenient in general practice and an indirect method, based upon the ionizing property of the radiation, is usually employed. In attempting to convert ionization values into absolute values, the question arises as to how much energy is expended in the formation of an ion-pair. This problem has attracted a large number of investigators whose conclusions have differed from time to time. It will therefore be of interest to review the evidence.

X rays and γ-rays do not ionize a gas directly but act through the medium of the secondary electrons ejected by them from the atoms of the gas, the initial velocities of the electrons depending on the wave-length of the incident radiation. We are therefore really concerned with the energy expended by a secondary electron (or secondary β-particle) in the production of an ion-pair. Since it is more direct to use either artificial electron beams or the primary β-radiation from radioactive substances, many investigators have attacked the problem in this way. There is also considerable evidence available from measurements with α-rays. In this *Report* we shall compare the results obtained with these various sources, adopting the usual symbol ϵ for the energy per ion-pair, when this is expressed in electron volts.

It is to be expected that ϵ would depend on the ionization potential of an atom and would therefore not have the same numerical value for all gases. As regards a particular gas, the important point is whether ϵ is a constant or is dependent upon the velocity of the ionizing corpuscle (α-particle, β-particle or secondary electron).

We will therefore confine our attention mainly to the measurements in air, since this is perhaps the most interesting medium of all.

Ionization by X rays. The early work with X rays, using beams rendered homogeneous by crystal reflection, indicated that ϵ was dependent on the wave-length of the radiation (that is, on the velocity of the ejected secondary electrons). Thus Boos[1] found a three-fold increase in the ionization energy from 0·56 to 0·325 A., while Grebe and Kriegesmann[2] and Kriegesmann[3] found ϵ in air to increase from 87 V. at 0·397 A. to 282 V. at 0·166 A. A year or so later, Kulenkampff[4] obtained entirely different results, showing that ϵ in air was 35 ± 5 V. and independent of wave-length in the region 0·56 to 2 A. He considered that the energy available in the crystal-reflected radiation previously used was too small to measure accurately and accordingly used the narrow spectral bands near the Kα lines of various elements, the radiation being made fairly homogeneous by means of filters. The energy of the X rays was degraded into heat in a thin silver screen and measured by an iron-bismuth thermocouple.

The constancy of ϵ over the range 0·5–1·5 A. was confirmed by Kircher and Schmitz[5], who used both air thermometer and bolometer for the energy measurements. For these wave-lengths ϵ was found to be about 21 ± 1 V., but gradually increased to 31 V. as the wave-length was decreased from 0·5 to 0·2 A.

Rump[6] using unfiltered and filtered X rays excited at voltages between 40 and 150 kV. first found ϵ to vary between 330 V. for long wave-lengths and about 40 V. for short wave-lengths. He threw doubt on his own method of estimating the effective wave-lengths of the radiation, particularly in the case of the unfiltered radiation and showed that, by slight adjustment of $\lambda_{\text{eff.}}$, ϵ could be considered to have an approximately constant value of about 42 V. In a later paper[7], after recalibrating his ionization chamber, he gave a corrected value of 33 V. for the range 0·12–0·43 A.

The experiments of Bouwers[8] on the photographic action of X rays have led to further evidence regarding the constancy of ϵ. Bouwers determined the relation between the photographic blackening B and the absorbed energy E for a variety of wave-lengths, measuring the energy by means of a bolometer. Berthold and Glocker[9], when comparing photographic blackening with ionization I at various wave-lengths, found that B/I changed with the wave-length in the same way as Bouwers had found B/E to change. They concluded that E/I was approximately independent of wave-length.

By means of a Geiger counter, Steenbeck[10] determined the number of quanta absorbed in a certain volume of gas and from an estimation of the fraction of ions entering the counter, obtained a measurement of the energy of the X-ray beam. This method gave ϵ to be 28 V. at 1·54 and 2·29 A., but as Steenbeck estimated the possible error as ± 6 V., the independence of wave-length was not very decisive.

Crowther and Bond[11] also found ϵ to be constant in the wave-length region 0·60–0·75 A., these being the effective wave-lengths of relatively wide spectral bands. Their energy measurements were carried out by passing the X rays through one of two similar calorimeters filled with oil and estimating the temperature

difference between the two calorimeters by means of copper-eureka thermocouples. They found ϵ to be $42{\cdot}5 \pm 0{\cdot}4$ V.

Gaertner[12] using the CuKα line (1·38 A.) gave ϵ as 36·4 V.

Thus while the numerical values obtained in the X-ray region by the various investigators differ considerably, there is a general suggestion that ϵ is approximately constant, at least within the limits of accuracy claimed by each observer. The mean value of the preceding results is about 33 V.

Ionization by γ-rays. Turning now to energy measurements of the γ-rays from radium B and radium C, the available evidence, although very limited, is interesting. The measurements are rendered difficult because the heating effect due to the γ-rays is small and tends to be masked by the heating effect due to the α- and β-rays from the radium and its products. In the experimental arrangement devised by Ellis and Wooster[13] the effect of the α- and β-rays was compensated. The radioactive source was placed in a hollow copper rod inside a cylinder composed of four equal sectors, one of aluminium, the diametrically opposite one of lead and the intervening ones of insulating material. The difference in the absorption of the γ-rays in the aluminium and lead sectors created a temperature difference between them which was measured by means of thermocouples. The α- and β-rays were completely absorbed in the copper rod, the heating of which affected the aluminium and lead sectors equally. Ellis and Wooster found that the heating due to the γ-rays from the radium B and radium C in equilibrium with 1 gm. of radium was 8·6 calories per hour.

More recent work has led to slightly higher values. Zlotowski[14] using an adiabatic calorimeter designed by Swietoslawski and Dorabialska[15] estimated the heating effect of the γ-rays from 1 gm. of radium in equilibrium to be $8{\cdot}9 \pm 0{\cdot}3$ calories per hour. In this determination allowance was made for the heating effect of the α- and β-rays. Later, Swietoslawski and Zlotowski[16] placed the radioactive material in a tube surrounded by a water jacket, the whole being enclosed in massive lead absorption screens. The heat developed by the α- and β-rays was carried away by a stream of water flowing through the jacket. With this arrangement Zlotowski[17] estimated that the total heating effect due to the α- and β-rays from 1 gm. of radium and its equilibrium products is 130·5 calories per hour. From a redetermination of the total heating due to the α-, β- and γ-radiation using his former method, he estimated the heating effect due to the γ-rays alone to be $9{\cdot}1 \pm 0{\cdot}15$ calories per hour per gm. of radium.

For the determination of the total number of ion-pairs per sec. emitted by 1 gm. of radium and its equilibrium products Zlotowski turned to the work of Gray[18]. The latter determined the number of ions in air cavities in a large block of aluminium surrounding the radioactive material. By integrating the ionization values found at various distances from the radium, Gray estimated the total number of ion-pairs as $2{\cdot}13 \times 10^8$ per sec. per gm. of radium. Combining this with his own recent value for the heating effect, Zlotowski obtained for ϵ a value of $31{\cdot}2 \pm 0{\cdot}5$ V. In view of the difficulties attendant on this work, this value is in remarkably good agreement with the mean value of ϵ obtained in the X-ray region.

Ionization by cathode rays. The difficulties encountered in the energy measurements in the X-ray region, which no doubt are mainly responsible for the discrepant values, do not arise when cathode rays are used. The electrons from a hot metal filament can be accelerated by a known electric field and their velocity and energy at the point of entering the ionization chamber can be calculated. By measuring the ionization current and the electron current, the average number of ion-pairs per electron can be calculated, leading to a value for the energy of formation of each ion-pair. This is the basic idea underlying all the work on cathode rays.

Lehmann and Osgood[19] found that, for electrons accelerated at potentials between 200 and 1000 V., the average energy per ion-pair was 45 V. They observed a linear relation between the number N of ions per electron and the initial energy V of the electron (in volts) which could be expressed by the formula:

$$N = 0{\cdot}0225\,(V - 17) \qquad \text{.....(1),}$$

where 17 is the ionization potential of air.

Schmitz[20] using slightly faster electrons (1–9 kV.) states that ϵ has a constant value of 42 V. but the curve given in his paper suggests rather that the value falls from about 46 V. at 1 kV. to about 40 V. at 6 kV. and then rises slightly to 42 V. at 9 kV. Over approximately the same range of velocities (4–13 kV.) Buchmann[21] found ϵ to be 31 ± 3 V. and independent of the electron velocity.

In a particularly careful research with high-speed cathode rays travelling at velocities between 9 and 59 kV., Eisl[22] found a value of $32{\cdot}2 \pm 0{\cdot}5$ V. for ϵ, independent of the electron velocity. This value was for some time accepted as the most reliable of all but we shall see that the work of later investigators throws doubt upon the constancy of ϵ, particularly for slow cathode rays.

Thomson[23] showed ϵ to vary from 55·5 to 38·3 V. when the initial electronic energy was changed from 54 to 265 V. He also observed a linear relation between the total number N of ion pairs per electron and the energy V of the electron and, expressing his results in a form similar to that of Lehmann and Osgood (equation 1) obtained the formula:

$$N = 0{\cdot}0270\,(V - 17) \qquad \text{.....(2).}$$

He points out that the nature of this formula is in agreement with theory, since when the electron voltage V is equal to the ionizing potential of the atom the number of ions formed must be zero. Assuming that the linearity was characteristic of electrons of all velocities, he expressed his results in the alternative form

$$\epsilon = \frac{37}{1 - 17/V} \qquad \text{.....(3),}$$

showing that as V approached ∞, ϵ approached a value of 37 V.

For electrons of energies between 100 and 300 V., Freund[24] found that on the average 30·5 V. were required for every ion-pair. For energies less than 100 V., ϵ was observed to increase steadily, reaching a value of 57 V. for electrons of mean energy about 10 V.

In Eisl's work, the potential across the ionization chamber was applied in such a way as should produce an additional acceleration of the cathode rays. Although he varied the chamber potential, Eisl was unable to detect any appreciable change of the electron velocities. Gerbes[25], using essentially the same apparatus as Eisl but with improved accuracy of measurement, has however demonstrated the influence of chamber potential and suggests that small corrections must be applied to Eisl's value of ϵ. By combining the corrected Eisl's values, over the range 9–59 kV., with Pigge's values[26] for nitrogen converted to air for the range 300–1000 V., Gerbes has arrived at an empirical formula which is valid for the range 300 V. to 60 kV.

$$\epsilon = 31{\cdot}62 + 5{\cdot}27/\sqrt{(V - V_i)} \pm 0{\cdot}08 \text{ V.} \qquad \ldots\ldots(4),$$

where V is the initial energy of the exciting electron in kV. and V_i is the ionizing potential of air in kV.

Recent workers with cathode rays are thus agreed that ϵ is dependent on the velocity of the electrons although, for velocities above 5 kV., the variation of ϵ is not pronounced.

Ionization by primary β-rays. There is very little evidence from the experiments with the primary β-rays from radioactive materials, no doubt because of the difficulties arising from the heterogeneous nature of the radiation. Wilson[27] measured the ionization due to β-rays of various velocities from Ra (B + C). These were sorted out magnetically and it was observed that kV^2 was sensibly constant for all values of the velocity V of the β-particle, k being the specific ionization or number of ion-pairs per cm. path of the β-particle in a gas at N.T.P. Since the energy lost by the particle is approximately proportional to the inverse square of the velocity, the approximate constancy of kV^2 suggests that the energy per ion-pair is constant for all β-ray velocities.

Ionization by α-rays. For the first 4 mm. of the range of the α-particles from Ra C′, Gurney[28] estimated that the average value of ϵ in air was about 33 V. Rutherford, Chadwick and Ellis[29] later gave an average value of ϵ of 35 V. for the whole range of an α-particle from Ra C′. Confining his attention to the question whether ϵ was the same for α-particles of lower velocity, Gurney showed that energies calculated from the energy-range curves of Marsden and Taylor[30] or Kapitza[31] agreed with the integrated areas under the ionization-range curves of Henderson[32]. He concluded therefore that the average energy spent per ion-pair was about the same at the end of the range of the α-particle as at the beginning.

Briggs[33] and also Curie and Joliot[34] arrived at the same conclusion. From his determination of the retardation of an α-particle from Ra C′, Briggs found that the curve showing the variation of the energy loss per cm. of path with the range was similar to the ionization-range curves of Henderson and of Curie and Behounek[35]. Curie and Joliot suggested that the energy per ion-pair was about 2 per cent less for the first 3 cm. of the range of an α-particle from Ra C′ than at the end of the range. The fraction of energy absorbed from an α-particle in the production of an ion-pair in air thus appears to be approximately constant but, as is pointed out by Rutherford, Chadwick and Ellis, this relation between ionization and energy

cannot hold for all gases, since the ionization curves for different gases are not always identical in shape. For example, while the ionization curves for helium and nitrogen are similar in shape to that for air, those for hydrogen and neon differ considerably from air. They conclude that the ionization is not in all cases proportional to the energy absorbed from the α-particle but depends on the velocity of the particle.

As regards the two values of ϵ for air which have been given for the α-particles from Ra C′, it might be mentioned that part of the ionization by these particles results from the electrons (or δ-rays) ejected by them from the atoms of the gas. The fastest δ-particles may have twice the velocity of the α-particles ejecting them, corresponding to an energy of about 4000 V. In general, the δ-particles will probably have speeds corresponding to energies of a few hundred volts. The values of ϵ for α-particles should therefore correspond to those for cathode rays having energies similar to the δ-particles.

The values of ϵ for the various ionizing sources are for convenience summarized in the following table. It will be seen that much of the data satisfies the empirical

Energy of formation of an ion-pair in air (ϵ)

Author	Effective wave-length of ionizing radiation	Average initial velocity of ionizing particles	ϵ in eV.
	X rays		
	A.	kV.	
Kulenkampff, 1926	0·56 to 2	22 to 6	35 ± 5
Kircher and Schmitz, 1926	0·21 to 0·5	60 to 25	31 to 21
	0·5 to 1·5	25 to 8	21 ± 1
Rump, 1927	0·12 to 0·43	100 to 30	33
Steenbeck, 1928	1·54, 2·29	8, 5·4	28 ± 6
Crowther and Bond, 1928	0·60 to 0·75	21 to 16	42·5 ± 0·4
Gaertner, 1929	1·38	9	36·4
	γ-rays from Ra (B + C)		
Zlotowski, 1935	0·006 to 0·05	2200 to 243	31·2 ± 0·5
	Cathode rays		
Lehmann and Osgood, 1927	—	1 to 0·2	45
Schmitz, 1928	—	9 to 1	42 to 46
Buchmann, 1928	—	13 to 4	31 ± 3
Eisl, 1929	—	59 to 9	32·2 ± 0·5
Thomson, 1931	—	∞	37
	—	0·27 to 0·05	38·3 to 55·5
Freund, 1935	—	0·3 to 0·1	30·5
	—	0·1 to 0·01	30·5 to 57
Gerbes, 1935	—	100 to 5	32 to 33·5 ± 0·08
	—	5 to 0·5	33·5 to 38 ± 0·08
	—	0·5 to 0·1	38 to 44 ± 0·08
	α rays		
	—	eV.	
Gurney, 1925	—	$7{\cdot}68 \times 10^6$	33
Rutherford, Chadwick and Ellis, 1930	—	$7{\cdot}68 \times 10^6$	35

formula (4) of Gerbes which, although derived from cathode-ray measurements, would thus appear to hold reasonably well for all the ionizing radiations.

References

(1) Boos. *Z. Phys.* **10**, 1 (1922).
(2) Grebe and Kriegesmann. *Z. Phys.* **28**, 91 (1924).
(3) Kriegesmann. *Z. Phys.* **32**, 542 (1925).
(4) Kulenkampff. *Ann. Phys.*, Lpz., **79**, 97 (1926).
(5) Kircher and Schmitz. *Z. Phys.* **36**, 484 (1926).
(6) Rump. *Z. Phys.* **43**, 254 (1927).
(7) Rump. *Z. Phys.* **44**, 396 (1927).
(8) Bouwers. Diss. Utrecht (1924).
(9) Berthold and Glocker. *Z. Phys.* **31**, 259 (1925).
(10) Steenbeck. *Ann. Phys.*, Lpz., **87**, 811 (1928).
(11) Crowther and Bond. *Phil. Mag.* **6**, 401 (1928).
(12) Gaertner. *Ann. Phys.*, Lpz., **2**, 94 (1929).
(13) Ellis and Wooster. *Phil. Mag.* **50**, 521 (1925).
(14) Zlotowski. *C.R. Acad. Sci.*, Paris, **199**, 284 (1934).
(15) Swietoslawski and Dorabialska. *C.R. Acad. Sci.*, Paris, **185**, 763 (1927).
(16) Swietoslawski and Zlotowski. *C.R. Acad. Sci.*, Paris, **200**, 660 (1935).
(17) Zlotowski. *J. Phys. Radium*, **6**, 242 (1935).
(18) Gray. See *Radiations from Radioactive Substances*, Rutherford, Chadwick and Ellis, p. 498 (1930).
(19) Lehmann and Osgood. *Proc. roy. Soc.* **115**, 609 (1927).
(20) Schmitz. *Phys. Z.* **29**, 846 (1928).
(21) Buchmann. *Ann. Phys.*, Lpz., **87**, 509 (1928).
(22) Eisl. *Ann. Phys.*, Lpz., **3**, 277 (1929).
(23) Thomson. *Proc. roy. Soc. Edin.* **51**, 127 (1931).
(24) Freund. *Ann. Phys.*, Lpz., **22**, 748 (1935).
(25) Gerbes. *Ann. Phys.*, Lpz., **23**, 648 (1935).
(26) Pigge. *Ann. Phys.*, Lpz., **20**, 233 (1934).
(27) Wilson. *Proc. roy. Soc.* **85**, 240 (1911).
(28) Gurney. *Proc. roy. Soc.* **107**, 332 (1925).
(29) Rutherford, Chadwick and Ellis. *Radiations from Radioactive Substances* (1930).
(30) Marsden and Taylor. *Proc. roy. Soc.* **88**, 443 (1913).
(31) Kapitza. *Proc. roy. Soc.* **102**, 48 (1922).
(32) Henderson. *Phil. Mag.* **42**, 538 (1921).
(33) Briggs. *Proc. roy. Soc.* **114**, 341 (1927).
(34) Curie and Joliot. *C.R. Acad. Sci.*, Paris, **187**, 43 (1928).
(35) Curie and Behounek. *J. Phys. Radium*, **7**, 125 (1926).

§ 6. ABSORPTION SPECTRA

By G. Shearer, M.A., D.Sc.

Our knowledge of the energy levels of the various atoms in the periodic table rests largely on the measurement and interpretation of X-ray emission and absorption spectra. From such measurements we know that there is a single K level, three L levels, five M levels, seven N levels, etc. and that the emission lines correspond to electron transitions between these levels, the possible transitions being governed by certain selection rules. The absorption edges correspond to electron transitions from inner levels to the outer regions of the atom.

As far back as 1919 Stenström discovered that the absorption edges were accompanied by a fine structure on the short wave-length side and in 1920 Bergengren and Lindh found that the position of an edge could vary with the state of chemical combination of the absorbing element. This fine structure consists of a series of secondary maxima and minima extending over a range of several hundred electron volts beyond the edge. Various theories were put forward to explain this structure and, for a time, it was believed that it arose as a result of multiple ionization. In 1931 Hanawalt showed that it became less marked as the temperature of the absorber increased and this fact rendered the multiple ionization theory untenable. Kronig[1] then showed that such a structure could have its origin in the crystal lattice of the absorber and this view has received considerable support from later work by Kronig[1], Coster and Veldkamp[2], Sandström[3], and others.

A short account of some of Sandström's experiments may serve to indicate some of the recent developments in connexion with X-ray absorption. Among the elements which he investigated a particularly interesting series is that extending from zirconium to tellurium. To make his results more easily followed the electron distribution in this region is shown in the following table.

Element	K	L_{I}	$L_{\mathrm{II}}L_{\mathrm{III}}$	M_{I}	$M_{\mathrm{II}}M_{\mathrm{III}}$	$M_{\mathrm{IV}}M_{\mathrm{V}}$	N_{I}	$N_{\mathrm{II}}N_{\mathrm{III}}$	$N_{\mathrm{IV}}N_{\mathrm{V}}$	$N_{\mathrm{VI}}N_{\mathrm{VII}}$	O_{I}	$O_{\mathrm{II}}O_{\mathrm{III}}$
36 Kr	2	2	6	2	6	10	2	6	—	—	—	—
37 Rb	2	2	6	2	6	10	2	6	—	—	1	—
38 Sr	2	2	6	2	6	10	2	6	—	—	2	—
39 —	2	2	6	2	6	10	2	6	1	—	2	—
40 Zr	2	2	6	2	6	10	2	6	2	—	2	—
41 Nb	2	2	6	2	6	10	2	6	4	—	1	—
42 Mo	2	2	6	2	6	10	2	6	5	—	1	—
43 —	2	2	6	2	6	10	2	6	6	—	1	—
44 Ru	2	2	6	2	6	10	2	6	7	—	1	—
45 Rh	2	2	6	2	6	10	2	6	8	—	1	—
46 Pd	2	2	6	2	6	10	2	6	10	—	—	—
47 Ag	2	2	6	2	6	10	2	6	10	—	1	—
48 Cd	2	2	6	2	6	10	2	6	10	—	2	—
49 In	2	2	6	2	6	10	2	6	10	—	2	1
50 Sn	2	2	6	2	6	10	2	6	10	—	2	2
51 Sb	2	2	6	2	6	10	2	6	10	—	2	3
52 Te	2	2	6	2	6	10	2	6	10	—	2	4
53 I	2	2	6	2	6	10	2	6	10	—	2	5
54 Xe	2	2	6	2	6	10	2	6	10	—	2	6

Sandström has measured the L absorption of many of the elements in this range and some of the conclusions deduced from these and other measurements can be briefly summarized. It appears to be a general rule that an absorption transition always takes place to the first incomplete level if such a transition is allowed by the ordinary selection rules which are known to govern emission lines. All three L edges have associated with them a fine structure extending for a considerable distance on the short wave-length side of the edge. In addition certain elements show a multiple structure in some edges. Thus the L_{II} and L_{III} edges of the elements 48–52 are triple, the members of the triplet being close together. The L_{I} edge is

single as also are the L_{II} and L_{III} edges of the elements 40–46. A similar multiple structure is found at other portions of the periodic table such as for the elements zinc to selenium.

When we examine the table we see that for the elements 48–54 the $O_{II}O_{III}$ level is in process of being filled. The selection rules permit of transitions from the L_I to the $O_{II}O_{III}$ level but not of transitions from L_{II} or L_{III} to this level. In the range 40–46 the $N_{IV}N_V$ level is being completed and the selection rules allow of transitions to this level from L_{II} and L_{III} but not from L_I. If absorption occurs to the first incomplete level, subject to the selection rules, we should expect the L_I edge for elements 48–52 to correspond to the transition $L_I \rightarrow O_{II}O_{III}$ and the L_{II} and L_{III} edges for elements 40–46 to correspond to transitions $L_{II} \rightarrow N_{IV}N_V$ and $L_{III} \rightarrow N_{IV}N_V$. These three transitions are those which correspond to the L emission lines γ_4, γ_1, and β_2, so that we should anticipate a very close agreement between the frequencies of the lines γ_4, γ_1, and β_2 and the absorption edges L_I, L_{II}, L_{III} respectively. Whenever the level in question is complete this agreement should cease to exist, since the absorption edge can no longer arise from a transition to this level. Sandström finds these predictions accurately verified. For example if the frequency difference between the L_{III} edge and the β_2 line is plotted against the atomic number, it is found that this difference is very nearly zero between 40 and 46 but rises continuously from 47 to 52. Similar results apply to the other edges and lines. These results go far to prove that absorption takes place to the first incomplete level subject to the selection rules being satisfied. There is some evidence also that the selection rules corresponding to quadrupole transitions do not apply to absorption spectra.

The above table of atomic levels throws light on the origin of the multiple L_{II}, L_{III} edges found for elements 48–52. For these elements the L_I edge which is single corresponds to a transition $L_I \rightarrow O_{II}O_{III}$ since the $O_{II}O_{III}$ level is still incomplete. The L_{II}, L_{III} edges must result from transitions to the $O_{IV}O_V$ level which is empty below element 54. Sandström finds that these latter edges are triple and that the member of the triplet which must be considered as the main edge is that having the lowest energy value. The most obvious explanation of the other members of the triplet is that they correspond to transitions to normally unoccupied levels associated with more than one atom, that is, associated with the crystal lattice. From elements 40–46 when the L_{II}, L_{III} edges arise from transitions to the incomplete $N_{IV}N_V$ level no such multiple structure is found. It occurs, however, in the elements zinc to selenium which have a corresponding position in the periodic table to that occupied by the group 48–53.

There appears to be little doubt that the fine structure or series of maxima and minima associated with the absorption edges arises from the crystal structure of the absorber. According to Kronig, the De Broglie wave associated with the electron involved in the absorption process undergoes scattering from neighbouring atoms and this, according to wave-mechanics theory, will give rise to such a series of maxima and minima as are observed. In the case of cubic crystals the distances of corresponding maxima and minima from the edge will vary inversely as the square

of the lattice constant. Coster has shown that there is good agreement between the theory and the observed results and that different types of crystal lattice give rise to correspondingly different fine structures.

References

(1) KRONIG. *Z. Phys.* **75**, 468 (1932, etc.).
(2) COSTER and VELDKAMP. *Physica*, II, **1**, 25 (1935).
(3) SANDSTRÖM. Dissertation, Upsala (1935).

§7. CRYSTAL STRUCTURE OF METALS AND INORGANIC SUBSTANCES

By W. A. WOOD, M.Sc.

From a practical aspect, the application of X-ray analysis to the study of alloys involves, mainly, the determination of the atomic arrangement for all concentrations of the constituents. Progress in this respect, on the lines of the early work of Westgren, has continued steadily, and structural data for a great number of the commoner alloy-systems are now available. The characteristic feature of an alloy, in equilibrium at a given temperature, is the formation of a definite structure which persists unchanged in type, altering only in the dimensions of the unit cell, as the composition is varied over a particular range; two such homogeneous phases are separated in general by a range of composition in which exists a mixture of both. The structure of a phase possesses therefore a certain elasticity whereby one kind of metallic atom may be changed for another, without upsetting the stable formation arising from the interaction of the various positive ions and the common system of valency electrons. The addition of fresh atoms of the solute to the structure of the solvent may take place in one of two ways: in the first, considering a binary alloy AB, we see that the atoms of A, if sufficiently small, may enter interstices found between the atoms of the parent structure A/B; in the second, the atoms of A may actually take the place of B atoms in the structure of the solvent A/B. In the case of the latter, substitutional solid solution, the replacement usually occurs at random, the different B atoms being equivalent. Under certain conditions, however, the atoms of A may replace only particular atoms of B; the A and B atoms then possess a definite distribution relative to each other in the composite lattice; the resulting structure, termed a superlattice, has recently received particular attention.

On the theoretical side, a study of the structures of alloys discloses the factors which are responsible for the type of structure assumed by a particular phase, and for its characteristic physical properties. This interpretation of alloy structures has advanced markedly in recent years. The more important points are surveyed below.

Role of atomic size. The entry of an atom A into the structure of the phase A/B will set up a local distortion of the lattice. The extent of the distortion must depend on the degree of dissimilarity of the two kinds of atoms. One aspect of such dissimilarity will be disparity in atomic size; this factor may be therefore expected, other things being equal, to govern the number of A atoms which can be taken up

by the composite lattice before instability is induced. In a recent study of the primary solid solutions in silver and copper, formed by addition of the elements of the *B* sub-group of the periodic table, Hume-Rothery, Mabbot and Channel-Evans[1] concluded that for favourable conditions of extended solid solution the diameter of the atoms of solute and solvent must not differ by more than 14–15 per cent. The atomic diameters in this instance refer to those given by the closest distance of approach of the atoms in the crystal structure of the respective elements. Interesting structures have been determined where the atomic volumes of the constituent elements differ widely, and where, in order to maintain a favourable size-factor, two of the smaller atoms have replaced one of the larger. Such an example is provided by Cu_2Mg[2], a complicated cubic structure with 24 atoms to the unit cell, which may be regarded however as a face-centred cubic structure, like that of copper, in which two copper atoms are replaced by one large magnesium; similar instances are given by Bi_2K, W_2Zr, Au_2Bi and (ZnNi)Mg. In these cases the atomic volume of the single smaller atom is of the order of half that of the larger. The same considerations apply to interstitial solid solutions. Here the elements involved are H, Bo, C and N. Haag has shown that when these atoms are alloyed with such metals as Fe, Ni, and Co, the structure of the resulting phases, and the range of composition over which they are stable, depend mainly on the relative size of the atoms and the size of the interstices available in the lattice of the metal. It is again possible, as in the structure of calcium carbide, for the smaller atoms to enter the interstices in pairs. The distortion of the parent lattice produced by these types of solution may be very pronounced. Thus, one edge of the face-centred cube of calcium is increased relatively to the other two when the carbide structure is formed and a tetragonal lattice results with an axial ratio of 1·16. Similarly the structure of martensitic steel, deviating from the body-centred cube of iron in consequence of the dissolved carbon, also appears to form a tetragonal lattice. It is interesting to note that this distortion is usually accompanied by a marked increase of hardness in the final structure.

Tendency to chemical combination. A further factor, pointed out by Hume-Rothery[3], is the tendency of elements to combine sharply into a stable compound rather than continuously in solid solution. Thus sulphur prefers to form stable sulphides; the solution of sulphur in the normal metals, and of similar elements in class VI B, selenium and tellurium, is thereby very restricted. The elements of group V B, phosphorus, arsenic, antimony and bismuth, being less strongly electronegative, may form slight solid solutions; thus arsenic dissolves in copper up to 6 atomic per cent. Those of group IV B, which are still less electronegative, may enter quite appreciably into solution, as in the case of silicon which is taken up by copper to the extent of 14 atomic per cent.

Structure of the pure components. The structure of each pure component, *A* and *B*, is characterized, according to the Bloch theory, by a definite electron system. If the structures are very different, then the addition of atoms of *A* to the structure of *B* will tend to produce marked distortion of the common electron configuration, and the range of solution will be restricted. Thus in the structure of silicon, each

atom has four valency electrons and is surrounded by four neighbours; by sharing one electron with each of the four neighbours, the atoms build up the characteristically stable octets of electrons. If now an atom such as copper, possessing only one valency electron, is introduced into the silicon structure, it is evident that the homopolar linkages cannot be completely maintained and that any extended solution is improbable. This is found to be the case.

The significance of the relation between electron configuration and atomic arrangement was implicit in the empirical rule given by Hume-Rothery (1926). Recently the question has been successfully approached theoretically by means of wave-mechanics and its preliminary solution represents a marked advance in the subject. The main features are summarized below.

Hume-Rothery's rule[3]. A phase which persists over a restricted range of composition may be regarded as an intermetallic compound in the sense that it corresponds to a restricted ratio of the constituent atoms. Hume-Rothery pointed out that the ratio of the total number of valency electrons to the total number of atoms in the unit cell often had the same value when the same type of crystal structure appeared in the phase-systems of different metals. A characteristic structure therefore, whatever the kind of the component atoms, was defined by a special electron-atom ratio. Thus the structure of the γ phase of Cu/Zn is a peculiar one containing 52 atoms to a cubic unit cell and possessing an atomic ratio of 4 ($Cu_5^{\cdot}Zn_8^{\cdot\cdot}$); the total number of valency electrons in the unit cell is $4[(5\times 1)+(8\times 2)]$; the electron-atom ratio is therefore 21 : 13. The same type of structure, and the same ratio of 21 : 13, is found in the alloy phases $Cu_9^{\cdot}Al_4^{\cdot\cdot\cdot}$, $Cu_9^{\cdot}Ga_4^{\cdot\cdot\cdot}$, $Cu_5^{\cdot}Cd_8^{\cdot\cdot}$, $Cu_{31}^{\cdot}Sn_8^{\cdot\cdot\cdot}$ and so on. An interesting set of examples is given by $Fe_5Zn_{21}^{\cdot\cdot}$, $Co_5Zn_{21}^{\cdot\cdot}$ and $Ni_5Zn_{21}^{\cdot\cdot}$ where, to satisfy the rule, the first atom must possess no valency electrons. Again a particular body-centred cubic structure is characterized by the ratio 3 : 2 of electrons to atoms. This structure is possessed by $Cu^{\cdot}Zn^{\cdot\cdot}$, $Cu^{\cdot}Be^{\cdot\cdot}$, $Ag^{\cdot}Zn^{\cdot\cdot}$, $Cu_3^{\cdot}Al^{\cdot\cdot\cdot}$, $FeAl^{\cdot\cdot\cdot}$ and $Cu_5^{\cdot}Sn^{\cdot\cdot\cdot\cdot}$. An atom-electron-ratio of 7 : 4 gives a close packed hexagonal structure, as exhibited by the phases $Cu^{\cdot}Zn_3^{\cdot\cdot}$, $Cu_3^{\cdot}Sn^{\cdot\cdot\cdot\cdot}$ and $Ag_3^{\cdot}Al_3^{\cdot\cdot\cdot}$.

Application of Bloch's theory to alloys[4]. It has become evident that the forces holding together the atoms of a metal represent a special type of linkage. The electrostatic attractions of oppositely charged ions, responsible for the equilibrium of the ionic compounds, such as Na^+Cl^-, are here inapplicable because the metallic ions are uniformly positive. Nor have attempts to apply to metals the conception of homopolar binding, arising from the mutual sharing of electrons between atoms, proved at all adequate. The application of wave-mechanics to the problem of the metallic structure, whilst still far from accounting for its stability, has however led to interesting results. It has been possible to explain a number of physical properties and, in particular, to elicit the meaning of Hume-Rothery's rule. This has been accomplished, in the work of Bloch, Brillouin and H. Jones, by studying the behaviour of the valency electrons in the periodic field produced by the regularly arranged positive ions of a given structure.

In the Bloch model the state of an electron is defined by the three components

of a vector k which is a measure of the average momentum. By the exclusion principle no two electrons may exist simultaneously in the same state, so that each electron in a system corresponds to a definite volume in the momentum or k space, and the addition of further electrons progressively adds to the volume in k space. If an electron is free, its energy is proportional to the square of the vector k specifying its state, and, as it moves into unoccupied states, its energy varies in a continuous way. If however the electron moves in a periodic field, as in a crystal, this continuous energy relation ceases to hold; instead, the k space is divided up by planes, across which takes place a discontinuous jump in energy. The disposition of the planes depends on the crystal; the planes are parallel to all possible crystallographic planes in the structure, and vary in distance from the origin in accordance with the spacing of the corresponding crystallographic planes. The k space is therefore divided into a succession of Brillouin zones. The energy change involved in traversing the bounding planes of a zone is related to the intensity with which the corresponding crystallographic plane would scatter a beam of electron waves, and therefore, also, to the intensity of scattering of an X-ray beam. The latter is known from data of X-ray analysis. When the zones are constructed for a given crystal structure, it is possible to calculate the number of permissible states which may be allocated to one unit cell, or to one atom, and to compare this number with the available valency electrons possessed by that cell. The fact that the structure of the crystal determines the nature of the electron states, and the number of states per atom, is the significance of Hume-Rothery's empirical rule.

In the γ structure of brass, Cu_5Zn_8, the first planes to reflect X rays with any intensity are the (411) and (330), which have the same spacing. These planes therefore define the first Brillouin zone. H. Jones shows that the planes enclose a polyhedron of 36 faces and that the volume of the polyhedron in k space corresponds to 90 states per unit cell. The unit cell contains 52 atoms and, as indicated previously, 84 valency electrons. The first Brillouin zone is therefore almost completely filled, and this, according to Jones, is the condition characterizing the formation of a γ lattice. Jones has also applied the theory to the ϵ and η structures of brass and accounted for the variation in size of the unit cell with composition and for the range over which it is stable. He has considered also the case of bismuth and certain bismuth alloys, and has deduced for the conductivity, the diamagnetism and magnetostriction, values in good agreement with observation. Dehlinger[5] has studied the β phase, using the same methods; he finds that the permissible number of electrons per atom in the first zone is 2, whereas the observed ratio is 3 : 2, so that part of the electrons must be in the second zone, leaving the first incomplete. Dehlinger attributes to this the instability of the β phase at ordinary temperatures. These results indicate that Jones' method of treating the properties of metals and alloys represents a marked advance in the explanation of those structures in which the electron configuration plays the chief part.

Superlattices. When the different kinds of atoms in an alloy assume regular positions on the common lattice, instead of being distributed at random, they form a *superlattice*. Thus in the copper-gold system, at a composition corresponding to

Cu_3Au, the unit cell is face-centred cubic; in the disordered arrangement the atoms are distributed indiscriminately; in the ordered arrangement of the superlattice, the gold atoms occupy the corners of the cube and the copper atoms the centres of the faces. In certain cases, the shape of the unit cell may differ somewhat with the onset of an orderly arrangement. The original X-ray studies of superstructures were made on the systems of copper with gold, palladium and platinum by Johanssen, Borelius and Linde[6]. Other examples are found in gold-manganese[7], iron-aluminium[8], and iron-aluminium-nickel, now extensively used for permanent magnets, the quality of which is intimately associated with the superlattice structure.

The mechanism of the transition from the ordered to the disordered state has been treated theoretically by Dehlinger[9], Bragg and Williams[10], Bethe[11], and others[12]. The problem is one of equilibrium between the tendency of the atoms to take up the preferred ordered arrangement, characterized by minimum potential energy, and the forces due to thermal agitations tending to produce disorder. A quantity S is defined which specifies the degree of order, and a quantity V which is the increase in potential energy of the crystal when an atom moves from an ordered to a disordered place in the lattice. A first condition relating V and S at a given temperature T is given at once by application of the Boltzmann distribution formula. A second condition, in the treatment by Bragg and Williams, is given by the assumption that S is proportional to V and varies only slowly with T (in Bethe's theory, V is assumed to depend only on the order of the neighbouring atoms and not on S). The intersection of the curves given by the two conditions gives the equilibrium degree of order at any temperature. The principal feature of the results is the existence of a critical temperature T_c. At higher temperatures the thermal agitations produce disorder amongst the atoms; on cooling slowly through T_c, a sudden increase in the degree of order sets in, and a release of energy occurs due to the disappearing disorder.

The practical importance of the formation of superlattices is the associated change in physical properties of the alloy. Recent detailed studies of the changes in electrical conductivity and specific heat have been made by Sykes and Evans[13] for the iron-aluminium and copper-gold systems. Their results are in general agreement with the theoretical treatments. In the Cu/Au system, these authors show that the mechanism of the transition to an ordered state on cooling the alloy involves the formation of small nuclei which have a relatively high degree or order. The nuclei grow, under constant temperature, to the size of the individual grain, but, if the cooling occurs too rapidly for equilibrium, intermediate states are obtained. The existence of the nuclei within the grains leads to a novel type of X-ray spectrum in which diffuse lines, proceeding from the small nuclei, and sharp lines, produced in the normal way, exist together on the same photograph. In its practical aspect, therefore, as well as the theoretical, the study of the superlattice has resulted in marked progress in the knowledge of the structure of alloys.

References

(1) Hume-Rothery, Mabbot and Channel-Evans. *Philos. Trans.* A, **233**, 1 (1934); A, **233**, 44 (1934); A, **233**, 87 (1934).
(2) Runquist, Arnfeld and Westgren. *Z. anorg. Chem.* **175**, 43 (1928).
Jurrianse, T. *Z. Krystallogr.* **90**, 322 (1935).
(3) Hume-Rothery. Monograph and Report Series, No. 1, Inst. of Metals (1926).
(4) Brillouin. *Die Quantenstatistik* (Berlin 1931).
Jones. *Proc. roy. Soc.* **144**, 225 (1934); **147**, 396 (1934).
(5) Dehlinger. *Z. Phys.* **94**, 231 (1935).
(6) Borelius, Johanssen and Linde. *Ann. Phys.*, Lpz., **86**, 291 (1928).
Johanssen and Linde. *Ann. Phys.*, Lpz., **1**, 1 (1936).
(7) Bumm and Dehlinger. *Metallwirtschaft*, **13**, 23 (1934).
(8) Bradley and Jay. *Proc. roy. Soc.* A, **136**, 210 (1932).
(9) Dehlinger. *Z. Phys.* **83**, 832 (1933).
(10) Bragg and Williams. *Proc. roy. Soc.* A, **145**, 699 (1934); **151**, 540 (1935).
(11) Bethe. *Proc. roy. Soc.* A, **150**, 552 (1935).
(12) Hume-Rothery and Powell, H. M. *Z. Krystallogr.* **91**, 23 (1935).
Pierls. *Proc. roy. Soc.* A, **154**, 207 (1936).
(13) Sykes and Evans. *J. Inst. Met.* **58** (1936).

§ 8. CRYSTAL STRUCTURE OF ORGANIC AND BIOLOGICAL SUBSTANCES

By J. Thewlis, M.Sc.

The X-ray examination of protein structure has received a great deal of attention during the past few years. It is accepted, chemically, that the proteins consist of amino-acids united to form huge molecules. The amino-acids are united by peptide links and X-ray investigation has shown that the molecules exist in the form of long chains packed together to form fibres.

The simplest protein of all is fibroin, the silk protein, and in this case it has been shown that the molecular chain is roughly parallel to the direction of the silk fibre[1]. The work of Astbury has shown that this picture is to some extent true of all natural protein fibres. He found however that the X-ray photograph of hair was not consistent with the picture of a straight molecular chain[2], but he also found that on stretching the hair he obtained a photograph which did show the characteristics corresponding to such a straight chain. He therefore concluded that the molecules of the hair protein, keratin, must normally exist in a folded state from which they can be pulled out straight. The process was found moreover to be reversible.

The X-ray results lead to the following picture of the structure of hair keratin[3]. In its extended state (β-keratin) it consists of long straight molecular chains of amino-acids packed together side by side to form a three-dimensional grid. Normally this grid is in a buckled condition (α-keratin) but in the presence of water the folds can be pulled out by mechanical force (i.e. the hair can be stretched) to form the long straight chains characteristic of β-keratin. This picture also accounts for the "setting" of hair which is accomplished by steaming it in a stretched state. Under these conditions the breakdown occurs of certain of the cross linkages of the

grid which are put under stress when the molecule is stretched. These broken linkages re-form in new unstressed positions and the refolding of the chains does not now take place. It is found however that if the broken cross linkages are not allowed to re-form, the modified keratin molecule so obtained may be made to contract to a length even shorter than that of α-keratin. This has been called "super-contraction".

All proteins appear to be based on long molecular chains of amino-acids of approximately the same dimensions. Thus in the case of fully extended chains it is found that the average length of the amino-acid residue is about 3·5 A., i.e. there is a periodicity of about 3·5 A. along the direction of the fibre. The thickness of the amino-acid averages about 4·5 A. and the width is about 10 A. Also, all the trans-formation phenomena in keratin have been observed with myosin, the muscle protein[4], and similar effects have been noted in the case of fibrin, obtained from centrifuged blood and examined in the coagulated state and in the form of congealed threads[3].

The case of muscle is interesting as it represents a naturally occurring state of supercontraction. From its X-ray photograph, myosin must be presumed to exist normally in the folded or α-state, hence contracted muscle, if its elasticity is due to the myosin chain-molecules, must consist of supercontracted myosin, corresponding to supercontracted keratin. It should therefore be possible to transform both myosin and muscle into the straight-chain configuration. Astbury and Dickinson[4] have succeeded in doing this and have moreover induced myosin to show super-contraction as in muscle itself.

According to Worschitz[6] the amino-acids in proteins build up a molecular lattice cell which in the case of muscle has a length of 22 A. along the fibre-axis. These molecular cells in turn build up micelles the length of which he has estimated to be rather over 2000 A., the breadth being about 75 A. The micelle would thus appear to be about 100 molecular cells long. That the protein structure is by no means overcrowded has been shown in an interesting fashion by Worschitz and von Hermann[7] who have found that the calcific deposit which occurs in arterio-sclerosis is either inter- or intra-cellular. They found that in calcified aorta wall the X-ray photograph of protein was hidden by a number of lines due to the calcium salt, but on removal of these salts chemically the protein pattern was again visible. They found on the other hand[8] that the protein structure could be destroyed by the deposition of salts from hypertonic media (common salt, homatropin, etc.). With increasing concentration of the medium the X-ray lines due to the medium increase in intensity whilst those of the protein disappear. On washing out the hypertonic salt with water the protein pattern does not reappear. The suggestion is made that chemical action has caused a collapse of the polypeptide chains.

So far we have discussed the structure of natural proteins, but it may be of interest to consider the results of an attempt to examine the structure of synthetic proteins, or rather of synthetic polypeptides. Meyer and Go[9] have examined a series of glycylglycines from di- to hepta-glycylglycine and have also investigated a polyglycylglycine of molecular weight about 2000. As the hydrogen of glycine is

substituted in natural fibres by a number of different groups, the X-ray photographs of the glycylglycines do not resemble those of the natural proteins. They suggest that the molecular chains are arranged in layers 4·2 A. apart. This is not very different from the value of 4·5 A. mentioned above as the average thickness of the amino-acids in the molecular chains of proteins.

From the above outline it is clear that much progress has been made in the elucidation of the nature of the protein structure. The essential point in the theory of this structure is that the proteins consist of long molecular chains of amino-acids. The dimensions of the amino-acid residues have been measured and the chains themselves have been found to exist in either extended or folded states. It is with some surprise therefore that we learn from Svedberg's ultra-centrifuge experiments[10] that the molecules of many soluble proteins consist of large globular units or combinations of units of molecular weight 35,000. Moreover this result has been confirmed by the X-ray examination of pepsin and insulin[11]. In both cases approximately spherical molecules of molecular weight about 37,000 are present but the details of the structures are of course different. Insulin possesses a rhombohedral unit cell which contains a single Svedberg unit of molecular weight 37,000. Pepsin also has a rhombohedral unit cell, of much larger size, however, and with a very different rhombohedral angle. This unit cell contains six Svedberg molecules. The exact arrangement and shape of these molecules is not known, but comparison with insulin suggests that they are approximately spherical with radius 20 A. This is also the approximate shape and size of the insulin molecule and, as Svedberg gives a radius of about 22 A. for his spherical units, the confirmation would appear to be complete.

In spite of this work, however, it is a fact that most X-ray photographs of proteins suggest that the molecules consist of polypeptide chains, and the reason for this appears to be that the proteins as usually examined are in a different state from their original one. Such a change of state has been demonstrated very well by Astbury in the case of the globular protein, excelsin. By keeping this in water he was able to obtain a single crystal X-ray pattern corresponding to a very large unit cell, but superimposed on this was a typical fibre-pattern. Before the exposure the crystal had been soluble but afterwards this was no longer the case. A record had evidently been obtained of the change from a soluble globular protein to an insoluble fibrous one. Such changes can be illustrated in other ways. When proteins are deliberately denatured, for example by heating, the X-ray photograph of the resulting substance has a marked resemblance to that of non-oriented β-keratin[12]. Also, by stretching fibres and films of denatured albumins and globulins, X-ray photographs have been obtained which are typical of oriented bundles of straight-chain polypeptides.

What then is the nature of the difference between a globular protein and the fibrous protein to which it so easily gives rise; in other words, what is the constitution of the globular proteins?

Two possibilities present themselves: either the fibrous protein arises from the globular by a mere uncoiling of the molecular chain, or the chain form is the result

of linear polymerization or condensation of the globular form. There is evidence for both of these possibilities and, according to Astbury[13], it seems likely that the change involves first the dissolution of intra-molecular co-valent linkages, then the uncoiling of a chain system, and finally the coagulation or so-called crystallization of the liberated chains into bundles analogous to those of β-keratin. He emphasizes however that although denaturation of a protein always leads to the formation of polypeptide chain-bundles, it does not follow that all such chain-bundles are necessarily denatured.

The question as to whether or not this process of denaturation is reversible appears to be an open one. Anson and Mirsky[14] are convinced that it is reversible, but the possibility of reversibility seems to depend on the way in which the globular proteins are built up[13]. If they consist merely of coiled chains which can be uncoiled reversibly and can be prevented when uncoiled from agglomerating into parallel bundles, then the process of denaturation may be said to be reversible. If on the other hand they are built up at hazard, then reversibility would seem to be unlikely.

We come now to the most recent advances in this subject, which are again largely due to Astbury. He has presented a picture of the structure of the globular proteins which does not appear to be open to any serious objections[13]. It is based on the observations of Gorter and his co-workers[15] that in the region of *p*H 1 the surface area of a protein monolayer corresponding to a weight of 1 mg. is of the order (extrapolated to zero pressure) of 1 m^2. Now if the protein monolayer is formed of parallel arrays of polypeptides arranged more or less in a straight chain configuration, with their side-chains dipping into the substrate, then the area per mg. should be about 0·795 m^2. The conclusion is that protein monolayers are formed by the liberation of polypeptide chains from an originally globular configuration, the process being similar to that of denaturation.

Now the area of a protein monolayer was shown by Gorter and Grendel[16] to correspond to that of a set of spheres of radius about 22 A., which is that noted above as having been found by Svedberg, and by Bernal and Crowfoot in their analyses of pepsin and insulin. The obvious inference therefore is that globular proteins are simply curved monolayers with the side-chains directed radially. If however the density of such a system is calculated it is found to be about 1·14 whereas that of the globular proteins, like that of the fibrous proteins, is about 1·26.

Various possible arrangements can be tried, but all except one fail for one reason or another. The one satisfactory solution is to assume that the globular protein unit is built up by a set of four discs of monolayer of diameter about 40 A., which are placed on top of each other at a distance apart of about 10 A. (the characteristic side-chain spacing found by X-ray analysis). This structure corresponds both in weight and dimensions to that demanded by Svedberg's units, the density is correct and the liberated monolayer would have the required surface area. The arrangement also fits in well with the X-ray data as far as they go. Moreover recent work by Wrinch[17] on the quantitative explanation of the α-β-keratin transformation leads in quite a different way to a similar arrangement.

It will thus be seen that there has been built up a definite picture of the structure of proteins, both globular and fibrous, and, although many details still require to be filled in, it would seem that this picture cannot be very far from the truth.

References

(1) MEYER, K. H. and MARK, H. *Der Aufbau der hochpolymeren organischen Naturstoffe* (1930).
(2) ASTBURY, W. T. *Fundamentals of Fibre Structure* (1933).
(3) ASTBURY, W. T. and WOODS, H. J. *Philos. Trans.* A, **232**, 333 (1933).
WOODS, H. J. *Nature*, Lond., **132**, 709 (1933).
ASTBURY, W. T. and SISSON, W. A. *Proc. roy. Soc.* A, **150**, 533 (1935).
(4) ASTBURY, W. T. and DICKINSON, S. *Nature*, Lond., **135**, 95, 765 (1935).
(5) WORSCHITZ, F. and HERMANN, J. VON. *Fortschr. Röntgenstr.* **52**, 80 (1935).
(6) WORSCHITZ, F. *Fortschr. Röntgenstr.* **51**, 81 (1935).
(7) WORSCHITZ, F. and HERMANN, J. VON. *Fortschr. Röntgenstr.* **50**, 307 (1934).
(8) WORSCHITZ, F. and HERMANN, J. VON. *Fortschr. Röntgenstr.* **50**, 178 (1934).
(9) MEYER, K. H. and GO, Y. *Helv. chim. Acta*, **17**, 1488 (1934).
(10) SVEDBERG, T. *Chem. Rev.* **14**, 1 (1934). Also papers in *Nature*, Lond., *J. Amer. chem. Soc.*, *Kolloidzschr.*, etc.
(11) BERNAL, J. D. and CROWFOOT, D. *Nature*, Lond., **133**, 794 (1934).
CROWFOOT, D. *Nature*, Lond., **135**, 591 (1935).
(12) ASTBURY, W. T. and LOMAX, R. *Nature*, Lond., **133**, 795 (1934); *J. chem. Soc.* p. 846 (1935).
(13) ASTBURY, W. T. *Nature*, Lond., **137**, 803 (1936).
(14) Papers in *J. gen. Physiol.* during past ten years.
(15) Papers in *Trans. Faraday Soc.*, *Proc. K. Akad. Wet. Amst.*, *Biochem. Z.*, *Biochem. J.*, *Amer. J. Dis. Child.* and *J. gen. Physiol.* during past ten years.
(16) GORTER, E. and GRENDEL, F. *Proc. K. Akad. Wet. Amst.* **32**, 770 (1929).
(17) WRINCH, D. M. *Nature*, Lond., **137**, 411 (1936).

§ 9. RADIUM

By W. E. T. PERRY, B.Sc.

The measurement of the small amounts of radioactive substances present in ordinary materials has been the subject of numerous investigations, particularly in connexion with the estimation of the age of rocks by the helium method. Evans[1] has described refinements in apparatus and technique for the accurate determination of minute quantities of radium (10^{-11} to 10^{-13} g.) by the emanation method. The method consists in transferring the radon associated with the radium in the specimen to an ionization chamber and comparing the ionization it produces with that of a known quantity of radon. Gaseous samples are transferred directly to the chamber. Liquid specimens are boiled, the evolution of radon being assisted by bubbling a stream of nitrogen through the specimen. Solid materials are dealt with by a direct fusion process. This is carried out in an evacuated graphite resistance furnace, in which the specimen is heated to a temperature of 2000° C. and boiled for several minutes. The radon, which is completely removed from the specimen by this process, is transferred to the chamber by means of a current of nitrogen. The radon available from very weak specimens produces an ionization

current which is only of the same order of magnitude as the background ionization due to cosmic rays, local γ-rays and radioactive contamination of the chamber walls. To minimize the effect of fluctuations in the background ionization, a differential method of measurement is employed, in which the electrometer records only the difference between the ionization currents in two identical chambers, one of which contains the radon from the specimen. The deflection of the electrometer is recorded photographically, quantities of radon as small as 10^{-13} curie being observable in a 1-hour observational period.

The same apparatus is applicable to thorium analyses, but, on account of the short life of thoron (54·5 sec.), it is necessary to maintain a continuous supply of thoron to the ionization chamber. This is effected by a nitrogen stream, the rate of flow being adjusted so that the disintegration of as large a fraction as possible of the available thoron takes place in the ionization chamber. Owing to the difficulties associated with the complete removal of thoron from solid specimens, other ionization and tube counter methods have been applied to thorium analyses. One method, described by Finney and Evans[2], consists in counting the α-particles emitted by a polished rock surface or by a powdered rock. The counter, which is unresponsive to γ-rays and cosmic rays, records between 20 and 200 α-particles per hour per 50 cm^2 surface area of specimen. The separation of the effects of the thorium series of elements from those of the uranium and actinium series is accomplished by observing the counting rate when aluminium foils of various thicknesses are placed between the source and the counter chamber. The measurements are interpreted by means of tables of numerical constants computed from considerations of α-particle ranges, combined, in some instances, with separate uranium determinations by the emanation method.

Quantities of radon of the order of several millicuries enclosed in glass capillary tubes termed "seeds" are used in large numbers in medical practice and their measurement is effected by the γ-ray method similar to that used for radium. Macdonald and Campbell[3] have described apparatus incorporating a two-stage valve amplifier by means of which such seeds can be measured at the rate of about three per minute. The seed is dropped from a height of about 30 in. into an ionization chamber, being brought to rest in the middle of the chamber by a trap. The rapidly increasing ionization current produces a corresponding change in the plate current of the second valve and this is transmitted to a galvanometer by a transformer. The use of the transformer prevents the slow variations in the plate current of the valve from affecting the galvanometer and so eliminates unsteadiness in the galvanometer zero. For quantities up to about 3 millicuries the maximum swing of the galvanometer is proportional to the radon content of the seed.

The γ-ray method is also used in the determination of the equivalent radium content (10^{-3} to 10^{-1} mg. of radium El) of residues and concentrates met with in radium-refining operations. The sample, enclosed in a sealed container, is placed inside a large ionization chamber so that as large a fraction as possible of the γ-radiation contributes to the ionization. An accuracy of ± 1 per cent is attainable in the measurements. The accuracy of the final result is however largely dependent

on the correction which has to be made for the absorption of the rays in the sample itself since, on account of the comparatively large bulk involved, this may amount to as much as 25 per cent. Fränz and Weiss[4] have carried out measurements on materials of known radium concentration enclosed in glass tubes of different diameters and also on mixtures having various densities and known radium contents. Their results enable corrections to be estimated from the density of the material and the diameter of the containing tube, the resulting radium content value being accurate to within ± 5 per cent.

During the year under review, the results of the redetermination of the half-life periods of several radioactive elements have been reported. In the case of radium[5], the method adopted consisted in separating the ionium from a sample of an old unaltered mineral and comparing the amounts of radium produced by the ionium in known periods with the amount of radium originally associated with it in the mineral. The mean result obtained with samples of broeggerite and uraninite is 1686 years. The generally accepted value is 1590 years, which is based on the figure $3{\cdot}7 \times 10^{10}$ for the number of α-particles emitted by radium. Schiedt[6] has determined the period of uranium by counting the α-particles emitted by samples of uranium oxide, his value $4{\cdot}22_8 \times 10^9$ years being slightly lower than that obtained by Kovarik and Adams[7] by a similar experiment. The value 5·02 days has been obtained[8] for the period of Ra E by determining the instants at which two Ra E sources separately and together produce the same ionization current.

New results of calorimetric investigations of the energy of the radiation emitted by radium have been published by Zlotowski[9]. The total heating effect of the radiation was measured by observing the rate of rise in temperature produced in a calorimeter containing spherical lead absorbers 0·2–6 cm. thick which surrounded the radium source. The absorbers were suspended in an evacuated enclosure which was immersed in a water bath, the bath being maintained at the same temperature as the absorbers to prevent loss of heat from the calorimeter. The heat contributed by the γ-rays was determined by means of a smaller calorimeter containing lead absorbers 0·2–1 cm. thick, the effect due to the α-and β-rays being eliminated by surrounding the source with a water-cooling system. Zlotowski's value for the total heating effect produced by the radiation from 1 g. of radium in equilibrium with its short-lived products is $139{\cdot}6 \pm 0{\cdot}3$ cal. per hour, that of the γ-rays alone being $9{\cdot}1 \pm 0{\cdot}15$ cal. per hour. The latter figure does not differ very much from the figure 9·4 calculated from data concerning the number of pairs of ions produced in air by the γ-rays from Ra (B + C) and the energy required to form a pair of ions[10]. Values of the ratio of the energies emitted in the form of γ-rays by Ra B and Ra C and of the energy of the β-rays from Ra (B + C) deduced by Zlotowski are in agreement with those derived from ionization and other methods.

A further member of the radium series, polonium, has also been the subject of calorimetric measurements. Saniélevici[11], using a microcalorimeter, finds that the energy emitted by 1 curie of polonium amounts to 27·24 cal. per hour.

A recent achievement is the preparation of protactinium (in the form of the oxide Pa_2O_5) in a quantity sufficient for an accurate determination of its atomic

weight to be made by chemical methods[12]. The initial material consisted of residues from radium-refining operations at Joachimstahl. The concentration of protactinium was first increased from 1 part in 3,000,000 to 1 part in 5000 by a large-scale extraction process carried out on 1200 kg. of the initial material[13], laboratory methods of refinement being employed for the final isolation of the pure oxide. About 0·1 g. of the oxide was used in the atomic weight determination, the value obtained being 231. This value is in agreement with Aston's evidence that actinium lead, the end product of the actinium series, has an atomic weight 207.

References

(1) Evans. *Rev. sci. Instrum.* **6**, 99 (1935).
(2) Finney and Evans. *Phys. Rev.* (2), **48**, 503 (1935).
(3) Macdonald and Campbell. *Rev. sci. Instrum.* **6**, 212 (1935).
(4) Fränz and Weiss. *Phys. Z.* **36**, 486 (1935).
(5) Gleditsch and Föyn. *Amer. J. Sci.* **29**, 253 (1935).
(6) Schiedt. *S.B. Akad. Wiss. Wien*, **144**, 2*a*, 5–6, 191 (1935).
(7) Kovarik and Adams. *Phys. Rev.* (2) **40**, 718 (1932).
(8) Pompei. *J. Phys. Radium*, **6**, 471 (1935).
(9) Zlotowski. *J. Phys. Radium*, **6**, 242 (1935).
(10) Rutherford, Chadwick and Ellis. *Radiations from Radioactive Substances*, pp. 161, 495 *et seq.* See also Binks, W., this *Report*, p. 347.
(11) Saniélevici. *J. Phys. Radium*, **6**, 73 (1935).
(12) Grosse. *Proc. roy. Soc.* A, **150**, 363 (1935).
(13) Grosse and Agruss. *Industr. and Engng Chem.* **27**, 422 (1935).

SPECTROSCOPY

Edited by W. JEVONS, D.Sc., Ph.D.

§ 1. ATOMIC SPECTRA

By M. F. CRAWFORD, Ph.D., McLennan Laboratory, University of Toronto

Multiplet structure. The analysis of any of the spectra, I, II, III, etc., of an element consists of two quite definite steps. The first step is purely an empirical ordering of the experimental data. It consists of establishing by the method of recurring wave-number differences a set of term values such that the wave-number of each line of the spectrum under investigation is given by the difference between two term values. The definiteness of such a term scheme is greatly increased by accurate values of the wave-numbers. The allowable tolerance as to the equality of the recurring differences is of the order of twice the uncertainty in the measured wave-numbers, and the smaller the tolerance the less probable the occurrence of differences due to chance.

Accurate wave-numbers alone will not always insure the exclusion of spurious differences in a spectrum rich in lines. A knowledge of the excitation characteristics of the lines is also desirable, since real differences should occur only between lines of the same or nearly the same excitation class. In general most sources emit lines originating in several stages of ionized as well as neutral atoms, whereas the analysis of, say, the II spectrum involves only those lines emitted by singly ionized atoms. By an experimental study of the variation of the relative intensities of the lines as the excitation is raised it is possible, however, to segregate the lines into classes corresponding to their origins in neutral, singly ionized, doubly ionized, etc., atoms. This information greatly reduces the number of lines to be considered in establishing a term system, and thus minimizes the occurrence of spurious differences and the arithmetical work entailed. The sources that have been used effectively to obtain controlled excitation are: temperature sources as developed by King, Paschen and Schüler tubes, and discharges activated by electron emission from hot cathodes for the I and II spectra; electrodeless discharges for the I to V spectra; and vacuum sparks with variable inductance and capacity for the higher ionizations.

Excitation data and better wave-length measurements in all spectral regions have appreciably improved the empirics of the more recently published analyses.

The second step in the analysis of a spectrum is the interpretation of the empirically established term scheme. This involves assigning J-values to the terms, identifying the term types, which is equivalent to specifying the manner in which the angular momentum vectors of the electrons couple to form the resultant J's, and identifying the electron configurations from which the terms arise. The main criteria for the correctness of the interpretation are that the combining properties, including relative intensities, and the distribution of the terms must be consistent

with theory, at least so far as theory is capable of predicting these features. The theory is comprehensive; even very complex configurations can be treated quantitatively. The theoretical methods cannot be discussed here, but two features that have been considered more critically in recent interpretations merit comment, namely coupling and configuration perturbations.

In specifying the term type of a J state either the nomenclature for Russell-Saunders coupling or that for jj coupling is generally used. A literal interpretation of the use of either of these nomenclatures means that the coupling is one or other of these two limiting types. Although the coupling for the lighter atoms generally approaches Russell-Saunders and that for the heavier atoms tends towards jj, yet in most cases it is strictly neither of these, but some intermediate type. The use of one of these nomenclatures to describe a spectrum, then, is justified provided it is understood that its use implies but a semi-quantitative description of the actual coupling, indicating that it only tends towards that limiting type to which the nomenclature is strictly appropriate. A quantitative expression for the actual coupling, such as is required for calculations on intensities, hyperfine structure, nuclear quadrupole moments and g-factors, can be determined from the term scheme only by an application of the quantum mechanical theory of multiplet structure[1].

In identifying the electron configurations from which the various states arise the usual procedure has been to compare the relative distribution and combining characteristics of the terms with those predicted for the various configurations when each is considered to be isolated and quite independent of the others. In many cases, especially for the light elements, the agreement between prediction and observation has been remarkably good; in fact it has been a vital factor in the rapid development of the comprehensive theory of spectra. The reason for the good agreement may be seen in the fact that for the lighter elements the separations between the J states of the same configuration are usually considerably smaller than the separations between the different configurations. The multiplet separations for the heavy elements, however, are frequently as large as the separations between the configurations, and often the structure is anomalous, in the sense that it is not that predicted for isolated configurations. Quantum mechanics shows that when the multiplet separations of two configurations of the same parity are comparable with the separation between the configurations, there is a mutual perturbation between those states of the two configurations that have the same J. The perturbation, which is inversely proportional to the separation between the states and frequently quite large, produces relative displacements of the terms which are evident as apparent anomalies in the multiplet structure. The recent recognition of these configuration perturbations has explained a number of perplexing irregularities in the spectra of both heavy and light elements. Thus in those cases where the terms from two configurations of the same parity intermingle it is rather meaningless to assign those J states that are perturbed to a single electron configuration since such states share their characteristics, none of them having the properties they would if unperturbed. In the language of quantum mechanics, each perturbed state is represented by a

certain linear combination of the unperturbed wave-functions of all the states of the two configurations that have the same J.

Specific examples of the points discussed are rather numerous to quote; they can be readily found in the current literature. In general they have been recognized more frequently in those spectra with measurable hyperfine structure, primarily because a quantitative interpretation of the hyperfine structure requires an accurate knowledge of all phases of the multiplet structure, including coupling and configuration perturbations. In fact hyperfine structure is an excellent criterion for the correctness of a multiplet analysis.

Hyperfine structure: nuclear moments. The determination of nuclear moments has progressed steadily since the inceptive investigation of Goudsmit and Back on the hyperfine structure of Bi I. The methods of determining nuclear moments have already been described, and the data available at the beginning of 1935 were summarized in table 1, p. 348, volume **2**, of these *Reports*. A supplementary table is presented here incorporating the revisions and extensions made during the past year. The data have been taken mainly from Bethe and Bacher's review on "Nuclear Physics"[2].

An accurate knowledge of the moments of the proton is highly desirable in view of the important role they assume in nuclear theory. The spin of the proton is definitely known to be 1/2. The magnitude but not the sign of its magnetic moment had been previously determined by the deflection of atomic and molecular beams in inhomogeneous magnetic fields. The value obtained, 2·9 nuclear magnetons, is about three times that expected on the assumption that Dirac's theory is applicable to protons. Further it was inferred, on the basis of Dirac's theory for a positively charged particle, that the magnetic moment pointed in the same direction as the angular momentum of the proton spin, i.e. that the magnetic moment was positive. The failure of the theory to give the correct magnitude of the magnetic moment threw doubt on the validity of the predicted sign. The sign has recently been determined[3] by an extension of the method of atomic beams. A beam of neutral atoms is first deflected in a weak inhomogeneous magnetic field; then by means of a slit certain magnetic components of the beam are selected and passed through a weak rapidly varying magnetic field to induce non-adiabatic transitions between the different magnetic levels. The emerging beam is then analysed by a strong inhomogeneous field arranged to deflect in a direction opposite to that of the first field. The character of the resultant pattern depends on the sign of the nuclear magnetic moment in such a manner that the sign can be uniquely determined[4]. Their results show that the magnetic moment of the proton is positive, as would be expected for a positively charged particle.

The nuclear moments of the deuteron also have been measured. Its spin is 1; the magnetic moment is positive and equal to 0·85 nuclear magneton[5].

The moments of the neutron are hardly accessible to a direct measurement, but they have been inferred from the moments of more complex nuclei. The data on nuclear moments lead to the important empirical rule that all nuclei of odd atomic weight appear to have half-integral spins, while all nuclei of even atomic weight

Table 1. *Nuclear moments: recent revisions and extensions*
(Supplementing table 1, *Reports*, **2**, 348 (1935))

Atomic number Z	Element	Mass number M	Nuclear spin I	Source H, D, B	Nuclear magnetic moment μ	Magnetic class	Observed moment ratio
1	H	1	1/2	D, B	2·9	I	
		2	1	B	0·85	III	
3	Li	6	1 c	D	~0·8	III	~0·25
		7	3/2	H, D, B	3·2	I	
19	K	39	3/2	H, D	0·40	I	1·81
		41	3/2 b	D	±0·22	I	
21	Sc	45	7/2 a	H	3·6	I	
30	Zn	67	3/2 b	H	−1·7	II	
31	Ga	69	3/2	H	2·1	I	0·79
		71	3/2	H	2·7	I	
33	As	75	3/2	H	1·5	I	
36	Kr	83	9/2 c	H	−1	II	
41	Cb (Nb)	93	9/2 b	H	4	I	
47	Ag	107	3/2 c	H	0·2	I	
		109	3/2 c	H	0·2	I	
48	Cd	111	1/2	H	−0·65	II	1·00
		113	1/2	H	−0·65	II	
50	Sn	117	1/2	H	−0·89	II	1·00
		119	1/2	H	−0·89	II	
51	Sb	121	5/2	H	3·7	I	1·32
		123	7/2	H	2·8	I	
54	Xe	129	1/2 a	H	−0·9	II	−1·11
		131	3/2 b	H	0·8	II	
56	Ba	135	3/2 a	H	0·9	II	
		137	3/2 a	H	0·9	II	
63	Eu	151	5/2	H	Larger	I	2·2
		153	5/2	H	Smaller	I	
72	Hf	177	3/2 c	H			
		179	3/2 c	H			
74	W	183	1/2 c	H			
77	Ir	191	1/2 c	H			
79	Au	197	3/2 c	H	0·3	I	
81	Tl	203	1/2	H	1·4	I	0·98
		205	1/2	H	1·4	I	
83	Bi	209	9/2	H	4·0	I	

Notes on table 1:

(i) Many nuclei with Z even, M even, have been observed to show no hyperfine structure, indicating either $I=0$ or that the magnetic moment is very small (*Reports*, **2**, 347 (1935)).

(ii) The I-values not followed by a small letter are definitely established. Those followed by a letter are probable values; decreasing probability is indicated by the letters a, b, c.

(iii) In the Source column, H means determined from hyperfine structure, B from band spectra, D from magnetic deflection of atomic beams.

(iv) Class I indicates those nuclei with Z odd, M odd; II those with Z even, M odd; III those with Z odd, M even.

have integral spins, of which many probably are zero. The total spin of a nucleus is the resultant of all the angular momenta of the orbital and spin motions of the nuclear particles, where the resultant is determined according to the rules of the vector model of quantum mechanics. Now the orbital angular momenta are always integral (in units $h/2\pi$). Thus the empirical rule connecting atomic weight and

nuclear spin must be interpreted as showing that both the proton and the neutron have half-integral spins. In the case of the proton it has been proven experimentally that the spin is exactly 1/2. Although the neutron spin is fixed from the preceding discussion only as half-integral, it is assigned the value 1/2. Simplicity is a strong argument in favour of this value since it seems to be a general rule that the spin is 1/2 for all the known elementary particles (viz. proton, neutron, electron—positive and negative), and also for the neutrino.

The magnetic moment of the neutron can be deduced most directly from the moments of the proton and deuteron. It can safely be assumed that the deuteron has no orbital angular momentum in its stable state; thus its observed spin, 1, must be attributed to the vector addition of the spins of the proton and the neutron in the deuteron. Thus the magnetic moment of the deuteron is the sum of the magnetic moments of the two particles. The magnetic moments of the proton and deuteron are both positive, that of the proton being larger than that of the deuteron. Thus the neutron has a negative magnetic moment equal to $0{\cdot}85 - 2{\cdot}9$, namely, $-2{\cdot}0$ nuclear magnetons. The sign is that which would be expected if the neutron had a negative charge; its magnitude is of the same order as that of the proton..

The representation of nuclear moments by vector models of the nucleus has been discussed in the previous *Report* (**2**, 350). The values of the magnetic moments of the proton and neutron used in these models are not consistent with the most directly determined values discussed above. In view of this and the fact that the models fail to account satisfactorily for the accurately known ratios of the magnetic moments of the isotopes of the same element[6,7], these models must be considered only as preliminary attempts to order nuclear moments.

Hyperfine structure: quadrupole moments. Deviations from the interval rule have been discussed and a number of examples quoted in the previous *Report* (**2**, 347, 349–50). The perturbations that cause deviations from the interval rule occur when two multiplet states are so close together that the separation between them is comparable with the hyperfine separations of the states. In these circumstances the F's remain valid quantum numbers, but the J's that normally would characterize the multiplet states lose their significance as quantum numbers and the relative distribution of the hyperfine levels is no longer consistent with the interval rule.

The deviations from the interval rule found by Schüler and Schmidt[8] for Eu, however, cannot be explained by perturbation effects, since the two isotopes ^{151}Eu and ^{153}Eu, both of which have a spin of 5/2, show different deviations. The deviations can be satisfactorily represented by adding to the ordinary expression for the interval rule an interaction term proportional to the square of the cosine of the angle between the nuclear spin and the electronic angular momentum. The addition of this term leads to the following expression for the relative positions of the hyperfine levels of a multiplet term:

$$E_F = E_0 + A.C + B.C\,(C+1) \qquad \text{.....(1),}$$

where

$$C = F\,(F+1) - I\,(I+1) - J\,(J+1) \qquad \text{.....(2)}$$

E_0 represents the energy (in cm.$^{-1}$) of the datum, which usually is chosen to coincide with the centre of gravity of the hyperfine levels; the second term is the expression for the interval rule in which A is the interval factor governing the magnitude of the magnetic interaction between the electrons and the nuclear spin; B is the coefficient of the quantum mechanical expression for $\cos^2(IJ)$.

To illustrate how well equation (1) represents the experimental data and also to indicate the magnitude of the deviations from the interval rule, the calculated and observed separations of the hyperfine levels of $^{10}P_{9/2}$ of $^{151, 153}$Eu are compared in table 2.

Table 2. *Hyperfine separations of* $^{10}P_{9/2}$ *of* Eu I

Constants: $A^{151} = 34{\cdot}56$, $B^{151} = -0{\cdot}0219$, $A^{153} = 15{\cdot}44$, $B^{153} = -0{\cdot}0456$ (unit = 0·001 cm.$^{-1}$).

Separation formulae derived from eq. (1)	^{151}Eu	^{153}Eu	Calculated	Measured	Magnetic
$F=7$ $7A+448B$	$F=7$, *0*		0	0	*0*
$=6$ $6A+72B$	$=6$, *232·1*		232	232	243
$=5$ $5A-160B$		$F=7$, *336·0*	336	336	*336*
$=4$ $4A-272B$	$=5$, 437·9	$=6$, 423·7	430	431	449
$=3$ $3A-288B$		$=5$, *513·0*	513	513	542
$=2$	$=4$, 614·2	$=4$, 597·5	606	606	623
		$=3$, 671·5	672	671	684
		$=2$, *731·1*	731	731	*731*
	$=3$, 758·4		758	758	764
	$=2$, *868·4*		868	868	*868*

The separation formulae listed in the first column apply to both isotopes. The four constants, A^{151}, A^{153}, B^{151}, B^{153}, were determined from four measured separations. The second and third columns contain the calculated positions (relative to $F=7$ level of ^{151}Eu) of the levels of the individual isotopes; the levels used to determine the constants are printed in italics. The fourth column contains the calculated positions for both isotopes together; in the cases where two levels are too close to be resolved their mean position has been calculated on the assumption that both isotopes are present in equal abundance. The fifth column gives the measured positions of the levels; and the sixth contains the positions, calculated by the interval rule, that are expected for a purely magnetic interaction. As can be seen, the agreement between the fourth and fifth columns is excellent, in contrast to that between the fifth and sixth. Thus equation (1), including a term in $\cos^2(IJ)$, represents the data for Eu I much better than the interval rule alone. It is to be noted that B^{153} is considerably larger than B^{151}.

Schüler and Schmidt pointed out that an interaction term proportional to $\cos^2(IJ)$ would be expected if the nucleus had an electric quadrupole moment. It is indeed reasonable that nuclei should have small quadrupole moments. Such a moment arises if the protons are not distributed throughout the nucleus on the average with spherical symmetry. Two possibilities occur for those nuclei that are not spherically symmetrical: the distribution of positive charge in the nucleus may be spread out more in a direction parallel to the angular momentum I, or spread

out more in a direction perpendicular to I. The former type is an elongated nucleus with a positive quadrupole moment, the latter a flattened nucleus with a negative quadrupole moment.

The effect of these two types of quadrupole moment on hyperfine structure can be qualitatively deduced quite easily. It is obvious that the effect will depend on the distribution of the extra-nuclear electronic charge for the multiplet state in question. Three cases are to be distinguished: First, when the electronic charge distribution is spherically symmetrical the presence of a quadrupole moment of either type will have no effect on the hyperfine structure separations, since the electrostatic interaction between the quadrupole moment and the electronic charge will be independent of the relative orientation of I and J. Secondly, when the electronic charge is spread out in a direction parallel to J, i.e. elongated, the electrostatic interaction will depend on the relative orientation of I and J, and will give for a positive and a negative quadrupole moment a negative and a positive value of B, respectively, where B is the constant in equation (1). Thirdly, when the electronic charge is spread out more in a direction perpendicular to J, i.e. flattened, its interaction with a positive and a negative quadrupole moment will give rise to a positive and a negative value of B, respectively. The direction and relative magnitudes of the displacements of the hyperfine levels from their positions as predicted by the interval rule can readily be determined for both positive and negative values of B by means of equation (1).

The electrostatic interaction of a nuclear quadrupole moment with the extra-nuclear electronic charge has been theoretically computed by Bethe* and Casimir[9]. Their calculations show that in general the presence of a quadrupole moment adds an energy term proportional to $\cos^2(IJ)$ to the interval expression for the magnetic interaction, giving equation (1) which fits the experimental data so well. The expression for the electrostatic interaction involves three factors: the first factor is a mean value over the charge distribution of the electrons for the J state in question, the second is a mean value over the charge distribution in the nucleus, and the third involves the quantum numbers. The first and third factors can be evaluated directly; thus the second factor, which is a measure of the asymmetry of the nucleus, can be evaluated from the experimental data. The theory indicates that there is no quadrupole effect for nuclei with $I = 1/2$. Bethe's calculation for the case of a single electron outside closed shells shows that the quadrupole effect is also zero for states with $J = 1/2$, i.e. for s and $p_{1/2}$ electrons. Other states will be expected to show effects which are roughly proportional to their multiplet doublet separations; thus the larger deviations from the interval rule should be observed for low $p_{3/2}$ and d electrons of the heavy elements. The case of several valence electrons is more complicated, but, as shown by Casimir, can be treated by setting up the appropriate zero-order wave-functions for the J states in question. The wave-functions must be those corresponding to the actual state of coupling, which can be determined analytically from the multiplet separations. Relativistic corrections also should be included in the calculation.

* Quoted in *Nuclear Physics*[2], p. 226.

The quadrupole moments of a number of nuclei have been evaluated[10]; they are:

Nucleus:	^{63}Cu	^{65}Cu	^{75}As	^{151}Eu	^{153}Eu	^{175}Cp	^{201}Hg	^{209}Bi
R^2 (cm^2):	$-0{\cdot}1$	$-0{\cdot}1$	$0{\cdot}2$	$1{\cdot}5$	$3{\cdot}2$	$6{\cdot}1$	$0{\cdot}5$	$-0{\cdot}5\times10^{-24}$

(R is the co-ordinate of a nuclear proton relative to the centre of gravity of the nucleus.)

Both positive and negative moments have been observed and they range in value from one to ten times the magnitude that would be expected if a single proton were unsymmetrically distributed in the nucleus.

In view of the deviations from the interval rule which result from the quadrupole effect, it can be considered safe to use the interval rule to determine I only for those states that show no effect. In one-electron spectra these states are the ones arising from only an s or a $p_{1/2}$ electron. As the interval rule cannot be applied to a state with $J=1/2$, it can now be used with safety only for certain many-electron states for which the electronic charge distribution is spherically symmetrical. The $^3S_1$ state arising from a configuration with two unpaired s electrons outside closed shells is ideal.

References

(1) Condon, E. U. and Shortley, G. H. *The Theory of Atomic Spectra* (Cambridge University Press and Macmillan, 1935).
(2) Bethe, H. A. and Bacher, R. F. "Nuclear Physics." *Rev. mod. Phys.* **8**, 82 (1936).
(3) Rabi, I. I., Kellogg, J. M. B. and Zacharias, J. R. *Phys. Rev.* **49**, 421 (1936).
(4) Rabi, I. I. *Phys. Rev.* **49**, 324 (1936).
(5) Kellogg, J. M. B., Rabi, I. I. and Zacharias, J. R. *Phys. Rev.* **49**, 641 (1936).
(6) Breit, G. and Rabi, I. I. *Phys. Rev.* **46**, 230 (1934).
(7) Crawford, M. F. and Bateson, S. *Canad. J. Res.* **10**, 693 (1934).
(8) Schüler, S. and Schmidt, T. *Z. Phys.* **94**, 457 (1935).
(9) Casimir, H. *Physica*, 's Grav., **2**, 719 (1935).
(10) Schüler, H. and Schmidt, T. *Z. Phys.* **98**, 430 (1936).

§ 2. ASTRONOMICAL SPECTROSCOPY

By A. Hunter, D.I.C., F.R.A.S., Imperial College of Science and Technology

Classification of stellar spectra. The spectra of stars are intimately related to so many other stellar characteristics that it is not surprising that the classification of stellar spectra should be regarded as fundamental in astrophysics.

As early as the middle of the nineteenth century, when enough data had been collected, it was realized, firstly, that the great majority of stars showed spectra consisting of absorption lines or bands crossing a background of continuous emission; and secondly, that these spectra could be arranged in a continuous sequence. Earlier classifications made on a basis of stellar colour thereupon gave way to Secchi's division of stellar spectra into four "types". The Harvard classification (the "Henry Draper Catalogue"), now in general use, employs the same principle, assigning to any given spectrum its place in the well-known sequence of types O, B, A, F, G, K, M, R, N, S according to the prominence or otherwise of certain selected groups of lines or bands.

The importance of this classification to astrophysical research can hardly be overestimated. Modifications have up to the present taken the form merely of refining the classification by what is in effect a system of decimal subdivision of the main types. For statistical investigations on low-dispersion spectrograms, this refined Harvard system, augmented by the "g" and "d" prefixes recognizing the distinction between giant and dwarf members respectively of a single type, will probably remain adequate for some time to come. The development of large telescopes, however, has made available high-dispersion spectrograms of the brighter stars, and this, in its turn, has recently raised the question[1, 2] of whether revision or even replacement of the Harvard classification is not desirable for the more detailed studies of stellar atmospheres now possible.

The trouble with the Draper classification results from its empirical basis: the characteristics distinguishing one type of stellar spectrum from the next are not fundamental physical ones. The spectrum of a star is, of course, conditioned by the various physical factors involved in the emission of its radiation; any one of these factors might conceivably form the basis of a system of classification. Systems using such criteria as effective temperatures and ionization in stellar atmospheres are, in fact, broadly in accord with the Harvard classification. Unfortunately, however, these parameters are not directly observable, nor would they form the most convenient basis for a practical classification. Conversely, criteria which *are* both directly observable and convenient, such as those used in the Draper catalogue, are not homogeneous, so that at different points in the stellar sequence the chief burden of discrimination between one type and the next falls first on one and then on another of the more fundamental characteristics.

The question thus lies largely between theory and practice. The theoretical astrophysicist requires a classification based on fundamental physical parameters; the practical astrophysicist one which expresses the gradation of the most conspicuous features of the spectra, not necessarily of the most fundamental ones. The theorist, however, is bound to admit that what he now regards as fundamental may not remain so for long. His theories, too, are still woefully inadequate to account for such phenomena as the distinction between the "carbon" and "nitrogen" series of Wolf-Rayet stars, and the fact that lines of the same multiplet may have different relative intensities in different stars. On the other hand, the practical man is only too conscious of the limitations of the Harvard system imposed by the multiplicity of the criteria it employs. Moreover, the present system takes no direct account of line profiles and central intensities, characteristics now well to the fore in astrophysical research.

Systems of classification based on theoretical parameters will not be without their disadvantages. There is reason to suppose, for example, that two stars with the same effective temperature and the same surface gravity would have very similar spectra, if the composition of their atmospheres were identical, and if certain other modifying factors were the same for both. Theoretically, then, we might hope that any stellar spectrum could be accurately specified by giving the appropriate numerical values to these two parameters. But there would still remain uncertainty

as to the atomic and molecular abundances and the turbulence in the atmosphere, the rotation speed, the interstellar absorption, and many other factors on which the final spectrum depends.

The conclusion to be drawn at present seems to be that although extensions of the Harvard classification are desirable for detailed studies of the atmospheres of bright stars, the greatest care must be used in making such extensions. In particular, revision must be regarded as premature in the absence of an exhaustive survey of representative spectra from λ2900 to λ10,000 A. on high-dispersion spectrograms; of accurate measurements of profiles, central intensities, and equivalent widths, made in conjunction with theoretical studies of the mechanism of line formation; and of further information on atomic and molecular abundances in the surface layers of stellar atmospheres.

As an example of the concentrated attention now being focused on removing the limitations of the Harvard classification may be mentioned a recent study[3] of F-type spectra. In this investigation a detailed quantitative examination is made of the standards now used for classifying stars in each of the subdivisions from F0 to G0. The observations involved the use of no fewer than 440 spectrograms of 201 stars.

The amount of information which can now be obtained from detailed examination of high-dispersion stellar spectrograms is well illustrated by a recent paper[4] on the "hot carbon" star R Coronae Borealis. This is an irregular variable whose peculiarities have led at various times to its assignment to types as widely different as F5 and G0. The following list of the physical parameters defining the state of this star's atmosphere will show in what astonishing detail we may hope to know the conditions existing in stellar atmospheres: composition, 69 per cent C, 27 per cent H, 4 per cent metals; effective temperature, 5300° K.; electron pressure at base of reversing layer, $8{\cdot}3 \times 10^{-7}$ atm.; mean total pressure, $9{\cdot}3 \times 10^{-6}$ atm.; level of ionization, 9·6 V.; surface gravity, $2{\cdot}7 \times 10^{2}$ cm./sec.^{2}; velocity of turbulence, 10 km./sec.; number of particles above 1 cm.^{2} of surface, $4{\cdot}9 \times 10^{21}$.

Planetary atmospheres. A large amount of visual, photographic, and spectroscopic work has been done in recent years on the problem of the constitution of planetary atmospheres. Each of these methods of attack has its advantages and disadvantages. Visual observations can detect temporary concealments of the surface markings where local condensations in a rare atmosphere are suspected, but can say little of the extent and nothing of the composition of the atmosphere. Direct photographs taken at intervals in light of various wave-lengths can prove invaluable for studying the opacity of and cloud-movements in planetary atmospheres, but help little in determining their chemical nature.

Spectroscopic observation would therefore appear at first glance to hold out the greatest hope both for the qualitative and for the quantitative analysis of the atmospheric gases. But the method suffers from inherent limitations. Chief amongst these is the presence of oxygen and ozone in the Earth's atmosphere. These combine to set the short-wave limit of atmospheric transmission at λ2900 A., thus making impossible the spectroscopic detection of such likely gases as He, Ne, A, H_2, N_2,

O_2, and CO by absorption of their resonance lines or bands. On the other hand, gases such as NO_2 and the halogens, which have strong absorption bands in the visible region of the spectrum, have not been detected on the planets, perhaps because of their chemical activity. The suspected common origin of all the planets suggests, too, that bands due to the Earth's atmosphere may mask those due to the atmospheres of other planets, unless these latter are very pronounced. This difficulty may, however, be partially removed by using the Doppler shift due to relative line-of-sight motion between the Earth and the planet; if this is large enough, it will cause a separation between telluric and planetary lines; if not, at least a distortion of profile, due to blending, may be looked for.

Classification of the planets by their spectra divides them into two main groups: those which show no absorption bands, or weak ones like the telluric bands in the solar spectrum (Mercury, Venus, Mars, Pluto, and the Earth); and those which have strong absorptions extending deep into the infra-red (Jupiter, Saturn, Uranus, and Neptune). It is worthy of note that the planets are divided in just this way by considerations of mass; and also that the atmospheres of the minor planets, the first group, contain oxygen compounds, whilst those of the major planets contain hydrogen compounds.

Mercury, owing to its small mass and correspondingly low velocity of escape, is not expected to have an extensive atmosphere. Visual observations[5] suggest the presence of dust clouds in some sort of atmosphere. Its composition, however, has not been found spectroscopically[6].

Venus has a reflecting power so high as to suggest that the planet is entirely cloud-covered. Colour-filter photographs show that the clouds are probably very high in the atmosphere, so that all our spectroscopic information refers to a thin layer of rare atmosphere above the clouds. This layer shows no trace of O_2 or water-vapour, but faint infra-red bands have now been definitely assigned[7] to CO_2.

Mars provides an interesting example of the limitations of the spectroscopic method. Direct observations of clouds and of the polar caps, as well as colour-filter photographs, leave no doubt that an atmosphere, fairly considerable in extent, does exist. Its constitution has not, however, been established spectroscopically. No evidence has been found for the presence of either O_2, H_2O or CO_2 in any considerable quantity, and it seems unlikely that the main constituent is any one of these.

Jupiter shows in the less refrangible parts of its spectrum strong absorption bands which have been traced[8,9] by direct comparison with laboratory spectra to the molecules CH_4 and NH_3. Early estimates[9] suggested that the atmosphere contains about 40 metre-atmospheres of methane and 10 m.-atm. of ammonia. More recent determinations[10], however, place the amounts at ten times as much as this. As late as 1932 it was thought that an absorption band at λ6191 A. was due to water-vapour, but a photometric study[11] of its contour showed conclusively that this identification was wrong; the band is now known to be due to methane. An interesting computation[12] of the temperature of the planet's visible surface from the partial pressure of NH_3 as determined spectroscopically puts the value at between

-100 and $-120°$ C. This agrees quite well with direct radiometric measurements, and confirms the theoretical deduction that the planet, if it were formed at the same time as the Earth, has had more than enough time to radiate away all its original store of heat, i.e. that its temperature is now determined solely by the amount of solar radiation it receives. At this temperature, of course, all water-vapour will be frozen out, and the visible clouds may be of CO_2. In neither case will the vapour pressure be enough to permit detection by spectroscopic means. As Jupiter has such a large velocity of escape, it is possible that a main constituent of its atmosphere is hydrogen; this, too, will not show itself spectroscopically.

Saturn's spectrum is similar to that of Jupiter, except that the NH_3 bands are weaker, and the CH_4 bands are stronger. Direct ultra-violet photographs show a layer of transparent atmosphere above dense clouds, and radiometric measurements suggest a lower temperature than Jupiter's. It seems probable, then, that in Saturn condensation of ammonia allows us to see more deeply into a methane atmosphere.

In Uranus and Neptune the NH_3 bands have disappeared, but those due to CH_4 are much stronger. Estimates[10] based on spectroscopic measurements place the amounts of methane in these outer planets at 4 and 25 mile-atm. respectively. The consequent strong absorption in the orange and red is responsible for the greenish blue appearance of these planets in the telescope.

Pluto probably has an atmosphere, but its constitution is at present unknown.

Instrumental advances. The spectrographic examination of faint stellar objects has up to the present been seriously limited by graininess in the fast emulsions necessary. Commercial interest in emulsions suitable for use in miniature cameras has, it is true, resulted in developer formulae designed to give finer grain, but such developers are usually non-contrasty and of low energy. Both these defects, of course, detract from their usefulness in astronomical spectrography. Recently, however, formulae have appeared for physico-chemical developers which are of high energy and give good contrast and a grain size only slightly larger than do the best of the less energetic developers. Experiments[13] at the Yerkes Observatory show that a developer using paraphenylenediamine, metol and glycin as developing agents produces, in spectrograms taken on the fastest commercial emulsions, detail and contrast comparable with those hitherto obtainable only with process emulsions. The saving in exposure time is, of course, enormous. Developers of this type seem likely to play an important part in expediting the routine work in almost all branches of observational astronomy.

In the 1934 *Report* (**1**, 129–30) it was pointed out that the extension of the technique of aluminizing glass surfaces to large telescope reflectors would open up the region $\lambda\lambda$3000–3500 A. to astronomical spectroscopists. The increased ultra-violet intensity obtainable with aluminized mirrors is particularly useful in photographing the spectra of late-type stars, whose intrinsic brightness is low in the ultra-violet. This region is of great importance because of the expected appearance there of many resonance and low energy-level lines, and many bands, such as the (0, 0) bands of the $^3\Pi \leftarrow {}^3\Sigma$ transition in NH, and of the $^2\Sigma \leftarrow {}^2\Pi$ transition in OH. A recent survey[14] of the spectrum of Arcturus (type K0) is typical of the progress made in

this branch of the subject. Strong lines of Cu, Cr and Al were observed, and the resonance lines of Ni, Ti and V are considered responsible for much of the general decrease in intensity of the spectrum in the ultra-violet. The second doublet of the principal series of Na at λ3302 occurs strongly, and it is pointed out that now we have the aid of aluminized mirrors, this pair of lines is particularly suitable for investigating the distribution of interstellar sodium. It is shown that the NH band at λ3360 is readily observable in spectra of this type, the only blending being with Ti^+ at λ3361. The importance of this band lies in the fact that from the intensity distribution in its rotational structure, the effective temperature of the reversing layer of the star can be easily calculated. The molecule NH is in all probability of common occurrence in the cooler stars, so that the band provides a useful criterion of stellar temperature and its variation with spectrum type. The OH band at λ3064, however, has not as yet been found in any stellar source other than the sun. It is unfortunately very close to the ozone absorption region, and is also expected to be badly blended with strong Ni and Cr multiplets.

In connexion with the aluminizing of telescope mirrors, it is interesting to note that the 200-in. disk destined for the new observatory on Mt Palomar, California, will be coated with aluminium.

Recent determinations of the radial velocities of the very remote spiral nebulae have been made possible only by the use of high-speed camera objectives in the spectrographs used. For instance, on the nebular spectrograph of the 100-in. telescope at Mt Wilson, an eightfold reduction in exposure times resulted from the substitution of an $f/0{\cdot}59$ lens[15] for the one previously used. With this lens the maximum aperture obtainable with a well-corrected "dry" lens had been reached. Further increase of numerical aperture, therefore, depended on the possibility of using an immersion-type objective. A lens of this type has now been made[16] with the extraordinary focal ratio of $f/0{\cdot}35$. The focus is on the plane surface of the rear component of the lens, which is in contact, via a suitable immersion fluid, with the photographic plate. The general design suggests an enlarged and slightly modified version of the ordinary 2 mm. immersion-type microscope objective. The use to which such a lens can be put is of course very limited; the spectrograph designed in conjunction with it was built solely to cover the range of spectrum (λλ3600–5000 A.) of special interest in work on the extra-galactic nebulae. Within these limits, however, the efficiency of the lens is bounded chiefly by the brightness of the Californian night sky.

It has long been known that intense magnetic fields—of the order of thousands of gausses—are associated with sun-spots. The existence of a general magnetic field at the Sun's surface, however, is not known with such certainty, since its magnitude is probably only of the order of 50 gausses. Recent progress in the design of measuring and recording machines, however, has made another survey[17] of the problem advisable. The magnitude of the task will be realized when it is mentioned that the Zeeman displacements of even those Fraunhofer lines selected as most suitable are only about 0·001 A. Repeated independent measurements of a large number of spectrograms has, however, confirmed the previous conclusions, and the existence of a very weak field may now be taken as established.

A survey of the present state of our knowledge of the spectra of the night sky and aurora was given in the 1935 *Report* (**2**, 339–40). It was there recorded, somewhat as a remarkable achievement, that a wide-aperture spectrograph had been successfully designed to photograph the spectrum of the night sky in 6 hours. A more recent paper[18] describes the construction of an instrument which will record the lines λλ5577, 5888 and 6300 A. with an exposure of only 15 sec.

References

(1) Struve, O. *Astrophys. J.* **78**, 73 (1933).
(2) Russell, H. N., Payne-Gaposchkin, C. H. and Menzel, D. H. *Astrophys. J.* **81**, 107 (1935).
(3) Hynek, J. A. *Astrophys. J.* **82**, 338 (1935).
(4) Berman, L. *Astrophys. J.* **81**, 369 (1935).
(5) Antoniadi, E. M. *J. Brit. astr. Assoc.* **45**, 236 (1935).
(6) Adams, W. S. and Dunham, T. *Publ. astr. Soc. Pacif.* **44**, 380 (1932).
(7) Adams, W. S. and Dunham, T. *Publ. astr. Soc. Pacif.* **44**, 243 (1932).
(8) Wildt, R. *Naturwissenschaften*, **20**, 857 (1932).
(9) Dunham, T. *Publ. astr. Soc. Pacif.* **45**, 42, 202 (1933).
(10) Adel, A. and Slipher, V. M. *Phys. Rev.* **47**, 787 (1935).
(11) Elvey, C. T. and Fairley, A. S. *Astrophys. J.* **75**, 373 (1932).
(12) Dunham, T. *Publ. astr. Soc. Pacif.* **46**, 231 (1934).
(13) Morgan, W. W. *Astrophys. J.* **83**, 254 (1936).
(14) Shaw, R. W. *Astrophys. J.* **83**, 225 (1936).
(15) Rayton, W. B. *Astrophys. J.* **72**, 59 (1930).
(16) Bracey, R. J. *Astrophys. J.* **83**, 179 (1936).
(17) Hale, G. E. *Nature*, Lond., **136**, 703 (1935).
(18) Garrigue, H. *C.R. Acad. Sci.*, Paris, **202**, 44 (1936).

§ 3. THE APPLICATION OF SPECTROSCOPY TO PHOTOCHEMISTRY

By R. W. B. Pearse, A.R.C.S., Ph.D., Imperial College of Science and Technology

Introduction. Systematic investigations of the chemical effects of light appear to have begun in the early part of the eighteenth century although some such effects had been known from very early times, for example, bleaching. That these photochemical effects, as they have been called, are of great interest and importance is realized at once when it is remembered that they have wide application in photography, that they play an important part in vision and that the growth of living things depends so largely on a satisfactory supply of light. Early work showed the sensitivity of silver chloride to light, the effect of light on chlorine water and on the rate of combination of chlorine and hydrogen, and incidentally led Ritter to the discovery of the existence of ultra-violet radiation. From this work emerged the first general law of photochemistry, the *law of Grotthus and Draper*, that only radiation which is absorbed is capable of producing chemical change. The later rapid development of the field, however, awaited the results of another rapidly growing field, that of spectroscopy, and the conception of the new ideas of the quantum theory.

According to the quantum theory radiation is emitted and absorbed in quanta, such that the energy of each quantum, E, is proportional to the frequency of the

radiation considered. This is usually expressed in the form of the well-known equation

$$E = h\nu,$$

where h is Planck's constant. Application of this idea to photochemistry by Einstein and Stark led to the second important generalization which is widely referred to as *Einstein's Law of Photochemical Equivalence*. In its most definite form this states that if a photosensitive system absorbs a quantity of radiant energy, Q, of frequency ν, then the number of molecules which react is given by $Q/h\nu$. This simple relation stimulated a vast amount of research in testing its validity for a great variety of reactions. Abundant evidence soon became available to show that exact equivalence held only for a small minority of cases, while for the majority, deviations occurred which were characteristic of the system in question. In some systems the number of reacting molecules was only a small fraction of the number of quanta absorbed; in others it was several thousands of times greater. It became evident that the majority of the reactions were complex in character and that exact equivalence must be restricted to the primary process of light absorption.

The primary absorption process. Consideration may first be given to the amounts of energy which are obtained by the system by absorption. If each molecule of a reactant receives one quantum, $h\nu$, the energy absorbed per gram molecule,

$$E = Nh\nu,$$

where N is Avogadro's constant. H. S. Taylor has compiled the following table (table 3) showing the magnitude of $h\nu$ in ergs and of E in calories for various regions of the spectrum. It is evident from this table that a system may gain very large amounts of energy by absorption of light. The further results of absorption will depend on whether the absorbing constituent is atomic or molecular.

Table 3. *Energy corresponding to various wave-lengths of light*

Colour of light	Wave-length	$h\nu$	E
Red	7500–6500	$2{\cdot}62$–$3{\cdot}02 \,.\, 10^{-12}$	37,800–43,630
Orange	6500–5900	$3{\cdot}02$–$3{\cdot}33$	43,630–48,060
Yellow	5900–5750	$3{\cdot}33$–$3{\cdot}42$	48,060–49,320
Green	5750–4900	$3{\cdot}42$–$4{\cdot}01$	49,320–57,880
Blue	4900–4550	$4{\cdot}01$–$4{\cdot}32$	57,880–62,330
Violet	4550–3950	$4{\cdot}32$–$4{\cdot}97$	62,330–71,800
Ultra-violet	2000	$9{\cdot}9$	142,000

Atoms consist of minute massive positively charged nuclei surrounded by numerous electrons. The energy of the atom depends on the configuration of the electrons. According to the quantum theory the electrons have a number of so-called "stationary states", each of definite energy. Change from one such state to another of lower energy is associated with emission of radiation while the reverse process is associated with absorption, according to the equation

$$h\nu = E' - E''.$$

All transitions between energy levels are not observed. Those that are allowed are easily obtained from a set of selection rules which have been derived. Spectroscopic investigations during the last half century have resulted in the observation of

the spectra of the majority of the elements of the Periodic Table and in the classification of their energy levels in terms of a consistent notation.

The absorption spectrum is in general much simpler than the emission because at ordinary temperatures all except a minute fraction of the atoms are in the lowest or normal state, consequently they can only absorb light corresponding to transitions, which are allowed by the selection rules from this state to higher states.

The atom, having absorbed a quantum of radiation and having consequently been raised to an excited state of greater energy, may lose this energy in one of several ways. It may fall back to its original level, re-radiating light of the same frequency, or it may fall back through intermediate levels radiating light of shorter wavelengths. This re-radiation of absorbed energy is observed as *fluorescence*. It has been studied extensively in the vapours of sodium and mercury by R. W. Wood[1]. Fluorescence is only obtained at low vapour pressures, a few millimetres of mercury giving the best results. Since the atom is known to remain in its excited state for an average time of only 10^{-8} sec. this indicates that collisions with other atoms are very effective in robbing it of its energy. This second mode of giving up its absorbed energy is, of course, the one which is the more interesting from the point of view of photochemistry. Many interesting experiments have been made in this field to study the quenching of the fluorescence[2] by admixture of various gases. In some cases by a small energy-change on collision the atom may be put into an energy state for which no transition to a lower by radiation is allowed. It is then said to be in a metastable state. The average life in this case may be as high as $\frac{1}{10}$ sec. The frequencies which are absorbed by an atom are extremely critical; if, for example, a mercury arc is used to excite mercury atoms care must be taken to ensure that the lines emitted by the arc are not self-reversed, otherwise very little effect will be obtained.

Molecules being structures composed of several atoms show correspondingly more complex spectra. The analysis of these is far from complete except in the case of diatomic molecules, although progress is being made with some polyatomic spectra. The quantum laws of emission and absorption apply to molecules as to atoms, but the energy of an electron state of a molecule has not one fixed value as for an atom but depends on the distances between the atoms composing the molecule. This dependence is represented in a very useful way by the curves introduced by Franck[3], which show the variation of the potential energy of a molecule with internuclear separation. In studies of diatomic molecules two types of curve have been distinguished, in the one, the "repulsive", type the potential rises continuously as the internuclear distance is decreased; in the other it first falls to a minimum, then rises sharply. The second type is necessary for the formation of a stable molecule. In its normal state at ordinary temperatures the molecule will be in the electronic level of lowest energy and near the minimum of the potential curve, which generally falls at an internuclear separation of about 10^{-8} cm. At the equilibrium distance the molecule may vibrate and rotate in various quantized states, the distribution of molecules among the states depending upon the temperature, in accordance with the Maxwell-Boltzmann relation.

By absorption the molecule will be raised from its normal electronic state to a higher one. The most probable transitions from a given vibrational level may be determined with the aid of the *Franck-Condon Principle*[4] which explains intensity distributions in band spectra. This is based on the assumptions that the majority of the electron transitions are likely to occur when the atoms are at one of the turning points of a vibration and that the transition will occur so quickly that the internuclear distance will not change appreciably. Thus on the Franck diagram the transition may be represented by a line drawn vertically, that is at constant internuclear separation, from the end-points of the vibration on the lower curve, to the upper. The theory is simple to apply in dealing with diatomic molecules and gives satisfactory results, but is complicated for polyatomic molecules and has not been applied very successfully so far.

The consequences of the transition will depend on the shapes and disposition of the potential energy curves. The molecule may dissociate or it may remain stable in the upper state until it loses its energy by radiation or collision. If the upper level is of the repulsive type the molecule will dissociate at once; if it is of the minimum type there are three possibilities. First, the molecule may be brought to a part of the upper curve where its potential energy is less than the value for infinite internuclear separation and may vibrate about the equilibrium position; secondly, it may be brought to a part where the potential energy is greater than this value and dissociate; thirdly, though brought to a stable part of the curve it may undergo a radiationless transition to another electronic state of similar energy where it is unstable and dissociate through this. The third possibility is referred to as *predissociation*.

In a given molecule it is possible to deduce from the appearance of its absorption spectrum which of the three results will follow the absorption of any particular wave-length. In a region where absorption leads directly to dissociation through a repulsive curve the spectrum is continuous. Where the molecule remains stable, absorption gives a band spectrum showing well-defined rotational structure. Towards shorter wave-lengths where the transitions are to higher points on the upper curve the bands sometimes become more and more crowded together in the spectrum and finally merge into a continuum. This marks the passage to parts of the curve where the potential energy is great enough for dissociation. Sometimes, however, the rotational structure of a band system becomes suddenly diffuse at a certain part of the spectrum. This marks the setting in of predissociation. When there is no indication of dissociation by diffuseness a useful estimate of the energy of dissociation can be made by extrapolation of a formula for the band-origins to the continuous region as was shown by Birge and Sponer[5].

Secondary processes. By absorption of monochromatic radiation an atom is raised to an excited state; by absorption of polychromatic radiation an assembly of such atoms may contain a number of different excited states. From the point of view of ability to undergo chemical reaction each of these may be looked upon as a new chemical substance. It is thus desirable to use monochromatic radiation in studying the resulting photochemical reactions.

The energy absorbed by the atoms which is not re-emitted in fluorescence may be lost during collision. Several types of process are possible. The energy may be used to excite other atoms electronically; Cario and Franck[6] found, for example, that the illumination of a mixture of mercury and thallium vapours by the resonance line of mercury led to the emission of characteristic thallium lines. Secondly, the energy may be utilized in chemical interaction with colliding gases; thus when hydrogen is mixed with mercury illuminated by the resonance line bands of mercury hydride appear in the emission spectrum. Again the energy may be used in causing dissociation of the colliding molecule; in this case the process is referred to as *photosensitization*.

Absorption by molecules may lead to more complex reactions since the molecule may be dissociated. The products of dissociation may then undergo further chemical reaction. This behaviour is well illustrated by the photochemical decomposition of hydrogen iodide, which has been the subject of much study. Hydrogen iodide shows a continuous absorption spectrum for wave-lengths less than about 3000 A. so that dissociation must occur. Determination of the quantum yield in the regions of 2070, 2530 and 2820 A. led in each case to a result of 2 moles per quantum within the experimental error[7]. The decomposition of two molecules instead of one by each quantum is explained by taking into account the subsequent reactions. Illumination gives the result

$$HI + h\nu \rightarrow H + I;$$

this is followed by the further reaction

$$H + HI \rightarrow H_2 + I,$$

and this again in due course by $I + I \rightarrow I_2$.

These three equations afford a ready explanation of the observed quantum yield. It should be noticed, however, that further reactions are conceivable which would tend to change this value. Thus decomposition of an iodine molecule by a hydrogen atom with the formation of a hydrogen iodide molecule

$$H + I_2 \rightarrow HI + I$$

would decrease the yield, whereas decomposition of a hydrogen iodide molecule by an iodine atom with the formation of an iodine molecule

$$I + HI \rightarrow H + I_2$$

would increase it.

Another reaction which may be compared with this is the combination of hydrogen and chlorine. Chlorine shows continuous absorption for wave-lengths shorter than 4785 A. The primary process will thus be dissociation of the chlorine molecule:

$$Cl_2 \rightarrow Cl + Cl.$$

In mixtures of hydrogen and chlorine two secondary reactions may follow:

$$H_2 + Cl \rightarrow HCl + H,$$

$$H + Cl_2 \rightarrow HCl + Cl.$$

Since each of these reactions produces a further free atom, a long chain of secondary processes might be expected. At atmospheric pressures yields of the order of 10^6 moles per quantum have been recorded by Bodenstein and Dux[8]. This is an example of what is called a chain reaction. Chain reactions of this sort are usually very sensitive to traces of impurities.

The decomposition of nitrogen dioxide shows results of considerable interest[9, 10]. At wave-lengths 5750, 5460 and 4360 A. zero yield was obtained. At 4050 A. the yield rose sharply to 0·74 and at 3650 A. and in the region 3160–2650 A. the yield was 2 within experimental error. The absorption spectrum shows a band system with sharp rotational structure from the red to about 3800 A. and diffuse bands from there down to about 2900 A. Norrish has examined the fluorescence spectrum and found that while 4360 and 4050 A. produced orange and greenish yellow fluorescences respectively there was an almost complete absence of fluorescence with 3650 A. A second region of absorption occurs with fine-line structure from 2695 to 2459 A. and diffuse bands to the further ultra-violet. The quantum yield in this region, however, does not appear to have been observed.

It is not, of course, possible in this short report to deal even briefly with the great number of photochemical reactions which have been investigated. A short bibliography of some of the more lengthy accounts which have appeared recently has therefore been appended[11–16].

References

(1) Wood, R. W. *Physical Optics* (Macmillan, 1934).
(2) Stuart, H. *Z. Phys.* **32**, 262 (1925).
(3) Franck, J. *Trans. Faraday Soc.* **21**, 536 (1926).
(4) Condon, E. U. *Phys. Rev.* **28**, 1182 (1926).
(5) Birge, R. T. and Sponer, H. *Phys. Rev.* **28**, 259 (1926).
(6) Cario, G. and Franck, J. *Z. Phys.* **11**, 161 (1922); **17**, 202 (1923).
(7) Warburg, E. *S.B. preuss. Akad. Wiss.* 300 (1918).
(8) Bodenstein, M. and Dux, W. *Z. phys. Chem.* **85**, 297 (1913).
(9) Dickinson, R. G. and Baxter, W. P. *J. Amer. chem. Soc.* **50**, 774 (1928).
(10) Norrish, R. G. W. *J. chem. Soc.* pp. 761–75 (1927); pp. 1158–69 (1929).
(11) Griffith, R. O. and McKeown, A. *Photo-processes in Gaseous and Liquid Systems* (Longmans, 1929).
(12) Style, D. W. G. *Photo-chemistry* (Methuen, 1930).
(13) Reprint and Circular Series of the National Research Council, No. 81, *First Report of the Committee on Photochemistry* (1928).
(14) Reprint and Circular Series of the National Research Council, No. 96, *Second Report of the Committee on Photochemistry* (1930).
(15) Noyes, W. A. Jr. *Rev. mod. Phys.* **5**, 280 (1933).
(16) Bodenstein, M. *Naturwissenschaften*, **23**, 10 (1935).

§ 4. THE ROTATION-VIBRATION SPECTRA OF THE LIGHTER HYDROCARBONS

By W. H. J. Childs, Ph.D., The Davy-Faraday Laboratory, The Royal Institution

It may fairly be said to-day that our knowledge of the structure of diatomic molecules is, in all but a few details, reasonably complete. The attention of the spectroscopist is now being transferred to polyatomic molecules, and it is not

surprising that the hydrocarbons, on account of their chemical importance and simplicity, should have attracted a rather large share. The information they have supplied is now considerable and a discussion of all the sources would be quite impossible within the length of this chapter. Since a selection of material is unavoidable, only the rotation-vibration spectra of some of the lighter hydrocarbons are dealt with and, as far as space permits, an account is given of some of the more interesting results which have recently been obtained. In the previous volume of these *Reports* (**2**, 1935) accounts were given of the general theory of the vibrations of polyatomic molecules (pp. 38–50) and earlier work on the infra-red spectra of the hydrocarbons (pp. 342–6). For a more complete treatment of the position up to the end of 1934 the reader is referred to a number of excellent monographs which have lately appeared[1,2,3].

Acetylene. Several factors combine to make the acetylene molecule especially favourable. Firstly, in common with most of the hydrocarbons, it possesses overtone bands of observable intensity in the region covered by the new infra-red sensitive photographic plates. This means that optical diffraction grating spectrographs of large dispersion and resolving power can be employed giving data of extreme precision. Again, the simple structure of the molecule is reflected in the simple rotational structure of the spectrum, so that its interpretation is obvious. Finally, it is relatively easy to calculate the fundamental frequencies for such a molecule on the assumption of simple force fields. The number of these frequencies is five for acetylene, and analysis of the spectrum soon showed the values of these and the probable modes of vibration to be associated with them. The two symmetrical vibrations, the one with the C atoms in phase with the H atoms, the other with the C atoms in opposite phase to the H atoms, have frequencies respectively 1975 and 3372 $cm.^{-1}$, whilst the antisymmetrical longitudinal vibration has a frequency 3288 $cm.^{-1}$ The two deformation frequencies are 600 $cm.^{-1}$ for the symmetrical and 730 $cm.^{-1}$ for the antisymmetrical. This allocation of frequencies satisfies the selection rules and is in agreement with the Raman spectrum; it also gives reasonable values for the C—H and C≡C bond constants. However, the values given are for molecules possessing one quantum of vibrational energy and are thus not suitable for comparison with the frequencies calculated from a model. In order to obtain the frequencies for *infinitely small amplitude*, which are the ones required, an extrapolation is necessary; this involves fitting the fundamentals and their overtones into a formula, and it is here that difficulties arise. From experience with the diatomic molecule we should expect that any vibrational energy level of the polyatomic molecule would be represented by an expression of the form

$$E = \Sigma\nu_i\,(v_i+\tfrac{1}{2}) + \Sigma x_i\nu_i\,(v_i+\tfrac{1}{2})^2 + \Sigma x_{ij}\nu_i\nu_j\,(v_i+\tfrac{1}{2})\,(v_j+\tfrac{1}{2}) + \ldots,$$

in which the coefficients x_i, x_{ij} result from the anharmonic nature of the vibrations. As a matter of fact the levels (we know now of 35 in all for acetylene) do not fit very satisfactorily in this type of expression; in some cases terms of higher order in $(v_i+\frac{1}{2})$ are necessary, and occasional unexplained discrepancies occur. This has led Mecke[4] to put forward recently quite another method of ordering the spectrum.

He concludes that where there are strongly anharmonic vibrations, and when the molecule is composed of several identical components (vibrations involving the C—H bond are always strongly anharmonic owing to the large amplitude of the H atom, caused by its lightness) the usual method of expressing the motion in terms of normal co-ordinates and their corresponding fundamental frequencies is quite unsuitable, though not, of course, wrong. Mecke suggests that it is better to regard the molecule as a set of coupled oscillators, and investigates the consequences of anharmonicity on such a system. The special advantage of this treatment is that the anharmonicity may be handled in its familiar diatomic form, and the troublesome transformation to normal co-ordinates is avoided. On this view the acetylene molecule is to be regarded as a set of seven coupled oscillators, one with a frequency of a C≡C component, two with a frequency of a C—H component, and four with the frequency with which a C≡C—H component would execute bending vibrations. Thus *three* instead of *five* frequencies are required to co-ordinate the spectrum, in conjunction with both coupling and anharmonic coefficients. The agreement between calculated and observed frequencies is much improved, and it becomes possible to derive a reliable value of 612 $cm.^{-1}$ for the infra-red inactive symmetrical deformation frequency. The frequencies of the three oscillators are found to be:

$$\nu_{CH} = 3428{\cdot}7 - 50{\cdot}5\,(v_{CH} + \tfrac{1}{2}),$$
$$\nu_{CC} = 1982 - 4{\cdot}5\,(v_{CC} + \tfrac{1}{2}),$$
$$\delta_{CCH} = 683{\cdot}7 - 6{\cdot}5\,(v_{\delta} + \tfrac{1}{2}).$$

A puzzling point with regard to acetylene has recently been cleared up by Funke[5]. The spectrum shows three distinct types of band: (*a*) with alternating intensities and no *Q* branch, (*b*) with alternating intensities and a fairly strong *Q* branch, (*c*) without alternating intensities and with a weak *Q* branch. Type-(*a*) bands are produced by longitudinal vibrations and type-(*b*) by the antisymmetrical deformation vibration. Type-(*c*) bands are all very weak and obscured by much stronger type-(*a*) bands so that their analysis is hardly possible, but it is assumed that they are produced when a deformation vibration is excited in both upper *and* lower states. This is quite likely, since the energies are so low that the excitation probability is appreciable even at room temperature. It has now been settled beyond all doubt by Funke, who has taken photographs with acetylene cooled to the temperature of solid CO_2; at this temperature the type-(*c*) bands disappear.

The deuterium substituted acetylenes are also of considerable interest. By a comparison of the moments of inertia of HCCH and HCCD it is possible to obtain absolute values of the internuclear distances. This has been done by Herzberg, Patat and Verleger[6], who thus obtain 1·057 A. for the C—H distance* and 1·204 A. for the C≡C distance. In such deuterium substitutions we have a powerful instrument for detecting flaws in our molecular models, for all we have done in the molecule (and all we have to do in our formulae) is to substitute a mass 2 for a mass 1. The deutero-acetylenes have, of course, not yet been examined with anything like the thoroughness expended on ordinary acetylene, but so far nothing has

* The C—H distance obtained in a similar way from HCN and DCN is also 1·057 A.

emerged to suggest that there is anything seriously wrong with the model we have been using. It is interesting to see how the substitution of one H atom by a D atom upsets the selection rules: the fundamental ν_2, which is inactive in HCCH, is of even greater intensity than ν_3 in HCCD. The values of the frequencies ν_2 and ν_3 are 3372 and 3288 cm^{-1} respectively for HCCH, and 3335 and 2550 cm^{-1} respectively for HCCD; all but the first of these four are active. Di-deutero-acetylene is proving obstinate. The heavier D atoms have a smaller amplitude than the H atoms they have displaced, which is unfortunate for the double reason that it means not only a smaller electric moment for the vibrations and thus a smaller intensity for the fundamentals, but also a smaller anharmonicity effect, which decreases the intensity of the overtones with respect to the fundamentals. In the photographic infra-red no bands due to C_2D_2 have yet been observed; in the far infra-red only one band, at 520 cm^{-1}, has been measured[7]; and in the Raman spectrum two frequencies have been measured[8], namely, 1762 and 2700 cm^{-1}

Methane. Next in importance to the acetylene spectrum is that of methane. Here the model is much more complicated, and the theoretical treatment correspondingly difficult. The molecule is presumably a regular tetrahedron, so that through symmetry the nine fundamentals are reduced to four, one single, one doubly degenerate, and two triply degenerate. The last are the only ones with an electric moment and are thus the only fundamentals to appear in the spectrum. The numerical values allocated to the fundamentals, in the order given above, are 2915, 1536, 1306 and 3020 cm^{-1}, and this agrees reasonably well with the far infra-red and the Raman spectra. In accounting for the overtones, however, the same difficulties are met with as in the case of acetylene. A number of attempts have been made without much success; it will be interesting to see whether the treatment suggested by Mecke[4] for acetylene can be applied to methane, and if so, what results it gives.

The methane spectrum presents the anomaly of bands with widely different rotational spacings, and for a time it was thought that the molecule could not be a regular tetrahedron. This difficulty has been settled by Teller[1], who has investigated theoretically the consequences of an interaction between vibration and rotation. In the case of a non-degenerate vibration the path of each nucleus is a straight line through its equilibrium position. When the vibration is degenerate this is no longer the case, and the motion is in general an orbit which does *not* pass through the equilibrium position. It is thus possible for the vibration to possess an internal moment of momentum.* The consequence of this is that each rotational level splits up into three (in the case of methane) with the vibrational moment respectively parallel to the moment of true rotation, zero and anti-parallel. The selection rules for the *P* and *R* branches allow transitions to the upper and lower, respectively, of each of these triple levels, so that the spacing in the bands is still regular, but *smaller* than that corresponding to the real moment of inertia of the molecule. The anomalous spacings (at least to first-order accuracy) are nicely explained by this theory, and it turns out that the internal moments for the two triply degenerate

* A good analogy would be that of one flywheel mounted upon another.

fundamentals are 0·45 and 0·05 (in $h/2\pi$ units), respectively. Knowing these values it is possible to derive an accurate moment of inertia. The value of this, $5{\cdot}267 \times 10^{-40}$ gm.cm^2, gives 1·09 A. for the C—H distance. The interaction has more recently been studied in greater detail and by more rigid methods by Jahn[9], who confirms Teller's results and extends them to include second-order effects. The effect on the spacings of overtone bands has been worked out by Johnston and Dennison[10].

The fundamentals first measured by Cooley[11] have been remeasured with apparatus of greater resolving power by Nielsen and Nielsen[12]; Steward and Nielsen[13] examined also the analogous bands of silane (SiH_4) and germane (GeH_4). Cooley's measurements are verified; but the interesting discovery is made that these methane bands possess a faint background structure, whilst the silane and germane bands are definitely (and unaccountably) complex. Nielsen suggests that the background structure of the methane bands may be due to transitions from excited *lower* levels, and measurements to test this are promised with methane at a low temperature. These will be awaited with interest, as fundamentals with sufficiently low energy to be excited appreciably at room temperature are very unexpected.

Measurements in the photographic infra-red have been extended from 1 to 1·2 μ by Childs[14]. This region includes the fairly strong bands at $\lambda\lambda$ 11,350, 11,620, 11,870 A., and the large dispersion and resolving power of the grating used reveal that the regularities which these bands exhibit when examined by prism-thermopile instruments are spurious. A feature of these bands, in common with those at wavelengths shorter than 1 μ, is their extremely complicated rotational structure; but a fainter band which is found at λ 11,050 A. is simpler, and this band is described and tentatively analysed.

The deuterium substitution is as powerful a weapon with methane as it is with acetylene. Unfortunately it is more difficult with methane than with acetylene to develop expressions giving the fundamentals in terms of the nuclear masses and the force constants (the straightforward classical calculation of normal co-ordinates and frequencies, simple in principle, is exceedingly laborious for anything but three or four masses). A good deal of information about the characteristics of the normal vibrations can, fortunately, be derived merely from a consideration of the symmetry of the potential field of a molecule, and in this respect group theory has proved a useful ally, but such a treatment can, of course, tell us nothing of the numerical value of the frequencies. Rosenthal[15] has shown recently that by assuming tetrahedral symmetry for the potential field of methane it is possible to express the normal frequencies in terms of five constants and the nuclear masses. As there are only four normal frequencies for methane, it is not possible to work backwards and determine the constants from the normal frequencies. It happens, however, that the internal moment of momentum which has been described above depends on the nature of the potential field, and if it is known it gives a fifth relationship which enables all five constants to be calculated. There is, in the case of one of them, an ambiguity of sign, but in actual practice this causes no difficulty. It is now possible,

by inserting the appropriate mass values, to calculate the normal frequencies of all four deutero-substituted methanes. The number of frequencies, of course, increases as the symmetry decreases, so that the four of methane become nine for di-deutero-methane, and eventually four again for tetra-deutero-methane. The calculated values (taking what appears to be the appropriate value for the ambiguous constant) are given in the accompanying table 4. The examination of the far infra-red spectra of the substituted methanes has not yet progressed very far[16] and the results are mostly of a preliminary nature, but those which have been published are given in the table for comparison. It will be seen that the agreement is not so close as to be uninteresting.

Table 4. *Fundamental frequencies calculated for the deutero-methanes*

(Values in parentheses are observed bands)

CH_4	CH_3D	CH_2D_2	CHD_3	CD_4
2914	2944 (2983)‖	2141	2101‖	2061
1520	1460 (1477)⊥	1317 1424	1286⊥	1075
3014	2183 (2205)‖ 3013 (3031)⊥	2227 2969 3013	2992‖ 2222⊥	2227
1304	1300 (1306)‖ 1151 (1156)⊥	1228 1019 1082	994‖ 1020⊥	987

In view of the relatively large quantities of gas involved it will be understood that the examination of the spectra of these methanes in the photographic infra-red is difficult. Childs and Jahn[17] have recently announced the successful preparation of 40 litres of mono-deutero-methane and the photographing of its spectrum to 1·2 μ. The absorption, as with ordinary methane, is particularly intense in the region 8350–9000 cm.$^{-1}$, and the rotational structure, though not regular, is simpler. There are ten fainter bands in the region 9600–10,200 cm.$^{-1}$ and four from 10,960–11,220 cm.$^{-1}$ Of special interest is a weak band with simple structure found at 9065 cm.$^{-1}$; in general appearance it is very similar to the methane band at 9047 cm.$^{-1}$, except that the approximate line spacing is 7·5 instead of 10·5 cm.$^{-1}$ This value corresponds to the moment of inertia about an axis perpendicular to the C—D axis, from which it would appear to be a parallel-type band; but the same spacing (taking into account the appropriate internal moment of momentum) is calculated for certain perpendicular-type bands to be expected in this neighbourhood, so that its allocation is uncertain until the rest of the bands are analysed.

Higher-membered hydrocarbons. A series of interesting measurements has recently been published by Gänswein and Mecke[18], who have examined the spectra in the photographic infra-red of the first six members of the paraffin series, five members of the olefine series and three members of the acetylene series. In this region we are confined practically to the third overtone of the C—H frequency 3000 cm.$^{-1}$

With the heavier members, of course, the rotation structure is too fine to be resolved; practically the limit of resolution of ordinary grating instruments appears to be reached with molecules containing only two carbon atoms. Most of the substances were examined as gases; the effect of condensation appears to be a fairly constant shift to longer wave-lengths of approximately 30–40 cm^{-1}, which is assumed to be due to the attractions of neighbouring molecules. Such comparisons as these bring out very well the interesting fact, which has excited attention since the beginning of infra-red work, that upon adding new radicles to a molecule and so making it more complicated, certain frequencies appear to be scarcely altered. These frequencies seem to be associated with certain groups of atoms, so that it is possible to label them as C—H, C=C frequencies for example. Why this should be so is not at all clear, as in general a molecule vibrates as a whole and not as a collection of parts; the fact that it is so invites us to an inspection of the theory. Mecke[19] goes into the problem in some detail, applying perturbation theory and the Schrödinger equation to a valency-force model and treating it as a collection of coupled oscillators. In the case of the hydrocarbons the results completely justify the allocation of frequencies to certain atom-groups, for the coupling and resonance factors turn out to be so small that the frequency shifts they cause are negligible. The consequence is that compounds containing a number of equivalent C—H valencies exhibit only *one* simple C—H fundamental, with an intensity proportional to the number of such valencies in the molecule. The actual frequency of this C—H fundamental depends on the degree of saturation of the C—C bonds; the third overtone, for example, is found at 10,930–11,170 cm^{-1} if modified by a C—C valency (paraffins), at 11,430–11,630 cm^{-1} if modified by a C=C (olefines), and at 12,660–12,740 cm^{-1} if modified by a C≡C (acetylenes).

The structure of the benzene molecule has recently been investigated in considerable detail by Ingold and his co-workers. This has been a subject of great interest since Placzek[20] published his work on the theory of Raman scattering, and deduced in particular the rule that molecules with a centre of symmetry cannot have fundamentals which appear both in the infra-red and Raman spectra. Since such coincidences had apparently been observed for benzene, it was concluded that the benzene molecule could not be a plane regular hexagon. A result in such flat contradiction to the chemical evidence needs thorough investigation. The infra-red, Raman, resonance and fluorescence spectra of both benzene and pure hexa-deutero-benzene have been examined by Ingold[21], who concludes that the evidence supports the D_{6h} model (regular plane hexagon) and that the coincidences between infra-red and Raman spectra, in so far as they are real, are such as might be expected with so complicated a molecule; they certainly do not rule out the regular plane hexagonal structure.

References

(1) Teller, E. *Hand.- und Jahrbuch der Chemischen Physik*, **9**, A 2 (1934).
(2) Mecke, R. *Hand.- und Jahrbuch der Chemischen Physik*, **9**, C (1934).
(3) Sutherland, G. B. B. M. *Infra-red and Raman Spectra* (Methuen, 1935).
(4) Mecke, R. and Ziegler, J. *Z. Phys.* **101**, 405 (1936).

(5) FUNKE, G. *Z. Phys.* **99**, 341 (1936).
(6) HERZBERG, G., PATAT, F. and VERLEGER, H. *Z. Phys.* **102**, 1 (1936).
(7) RANDALL, H. M. and BARKER, E. F. *Phys. Rev.* **45**, 124 (1935).
(8) GLOCKLER, G. and MORRELL, C. E. *Phys. Rev.* **45**, 569 (1935).
(9) JAHN, H. A. *Ann. Phys.*, Paris, **23**, 529 (1935).
(10) JOHNSTON, M. and DENNISON, D. M. *Phys. Rev.* **48**, 868 (1935).
(11) COOLEY, J. P. *Astrophys. J.* **62**, 73 (1925).
(12) NIELSEN, A. H. and NIELSEN, H. H. *Phys. Rev.* **48**, 864 (1935).
(13) STEWARD, W. B. and NIELSEN, H. H. *Phys. Rev.* **47**, 828 (1935); **48**, 861 (1935).
(14) CHILDS, W. H. J. *Proc. roy. Soc.* **153**, 555 (1936).
(15) ROSENTHAL, J. E. *Phys. Rev.* **45**, 538 (1934).
(16) GINSBURG, N. and BARKER, E. F. *J. chem. Phys.* **3**, 668 (1935).
(17) CHILDS, W. H. J. and JAHN, H. A. *Nature*, Lond., **138**, 285 (1936).
(18) GÄNSWEIN, P. and MECKE, R. *Z. Phys.* **99**, 189 (1936).
(19) MECKE, R. *Z. Phys.* **99**, 217 (1936).
(20) PLACZEK, G. *Leipziger Vorträger* (1931); English translation *The Structure of Molecules* (Blackie, 1932).
(21) INGOLD, C. K. *J. chem. Soc.* p. 211 (1936).

CAMBRIDGE: PRINTED BY WALTER LEWIS, M.A., AT THE UNIVERSITY PRESS